



저작자표시-비영리-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

도시계획학박사학위논문

대용량 자료를 이용한 네트워크 기반
도시간 경로통행시간 예측

Network-based Intercity Path Travel Time Forecasting
Using Large-scale Data

2015년 2월

서울대학교 환경대학원

환경계획학과

장 현 호

대용량 자료를 이용한 네트워크 기반
도시간 경로통행시간 예측

지도교수 이 영 인

이 논문을 도시계획학 박사학위논문으로 제출함

2014년 10월

서울대학교 환경대학원

환경계획학과

장 현 호

장 현 호의 박사학위논문을 인준함

2015년 1월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)



이 논문이 있기까지 시작과 끝을 함께해준
많은 분들께 감사드리며,
오랜 시간을 기다려준
사랑하는 아내(KJ)와 아들(JC)에게
KJC예측기의 원형인 이 논문을 바칩니다.

국 문 초 록

오늘날의 도시교통정보를 한 마디로 요약하자면 ‘대용량 자료’일 것이다. 첨단기술을 기반으로 하는 스마트 폰과 21세기 지능형교통체계(Intelligent Transport System, 이하 ITS)의 정보수집단을 통해 수집되는 방대한 양의 다양한 교통정보는 대용량 자료의 많은 부분을 점유하고 있으며, ITS 분야의 자료환경을 소량의 실시간 자료에서 방대한 이력자료를 포함하는 대용량 자료로 변화시켰다. 이러한 자료환경의 변화에 따라 최근 ITS분야에서는 대용량 자료를 수집·관리·분석하기 위하여 첨단자료관리시스템(Advanced Data Management System, 이하 ADMS)이 도입되고 있다.

ITS의 사회적 편익은 장래 교통상태의 예측을 통한 사전 교통류 관리 및 동적 교통정보제공으로 극대화된다. 따라서 교통상태의 예측은 ITS의 주요 요소 중 하나이며, ITS에서 교통상태 예측시스템은 첨단교통관리시스템과 첨단교통정보시스템의 핵심 하위시스템 중 하나이다. 이와 같이 교통상태 예측기술은 교통상태 예측시스템의 성능과 더불어 ITS의 편익 증대에 있어 밀접한 관련이 있다. 따라서 20세기 말에 ITS가 도입된 이후로 ITS 분야의 예측기술은 다양한 예측모형의 개발을 통해 지속적으로 발전하고 있으며, 하나의 학문분야로 자리잡았다.

기존의 고도화된 예측모형은 예측 정확도 향상이라는 목표를 달성하였으나, 다음의 한계를 가지고 있다. 첫째, 기존모형은 ITS 예측분야의 고질적 문제인 장래 상태의 불확실성을 극복하지 못 하였기 때문에 단기예측의 수준에서 벗어나지 못 하고 있다. 둘째, 많은 경우에 있어 실시간 자료를 이용하도록 설계되었기 때문에 ADMS와 같은 자료관리시스템에 탑재되어 실시간 자료와 대용량 이력자료를 이용하여 교통상태의 예측에 적용하기 어려운 구조적 문제가 있다. 마지막으로 고도화된 모형들은 ITS 시스템에 탑재·운영시 새로운 문제를 발생시키고 있다. 고도화된 모형은 모형의 구조변경, 입·출력 자료의 변경, 파라미터 값의 재정산 등에 교통류의 행태와 예측 모델링에 대한 깊은 이해를 필요로 하기 때문에 예측 모델링 경험이 부족한 현장의 운영요원에게 새로운 장애가 되고 있다.

본 연구에서는 ITS 예측분야의 지속적 도전 과제인 장래 상태의 불확실성을 극복하고 예측영역의 확장을 위하여 실시간 및 대용량 이력 교통자료를 이용한 교통정보 예측기(Forecaster)인 KJC 예측기를 개발하였으며, 다음의 목표를 달성하도록 설계되었다. 첫째, 장래 상태의 불확실성을 극복하기 위한 방안이 고려되었다. 불확실성을 감소시키기 위하여 입력자료의 공간적 영역을 기존의 지점 또는 구간에서 도로망으로 확장하고, 도로망의 소통상태를 이용하여 장래 상태의 불확실성을 감소시켰다. 둘째, KJC 예측기는 ADMS와 같은 자료관리시스템에 탑재되어 도로망 소통상태를 예측하고, 예측된 도로망 소통상태를 이용하여 단·중·장거리 경로통행시간을 예측하도록 개발되었다. 따라서 보다 적극적이고 전문적인 첨단교통관리와 동적 중·장거리 통행시간 정보제공에 활용할 수 있도록 하였다. 이상의 목적으로 개발된 KJC 예측기는 이력자료에 내재된 장래 교통상태 정보를 탐색 및 구축하기 위한 지식탐색 모듈, 군집화모형을 이용한 의사결정 그룹화 모듈, 그리고 사례기반 추론을 기반으로 하는 예측 의사결정 모듈로 구성된다. 3개 모듈은 입·출력 자료구조의 용이한 변경, 결측자료의 자동처리, 파라미터 값의 자동정산, 연산수행속도 등을 고려하여 개발되었다.

본 연구에서 개발된 KJC 예측기의 성능은 대용량 자료환경에서 평가되었다. 경부고속도로 서울-대전 구간을 대상으로 8개월간 약 4억건의 통행사슬 자료를 이용하여 도로구간 소통상태 및 경로통행시간 이력자료를 구축하였으며, 구축된 자료는 총 18,768,960건이다. 다각적인 종합평가 결과, 개발된 교통정보 예측기는 매우 빠른 연산수행속도를 보이면서 장래 6시간까지 도로망의 소통상태를 합리적으로 예측하였다. 그리고 개발된 예측기로 추정된 도시간 경로통행시간의 정확도는 모든 단·중·장거리 통행시간 시나리오에서 기존의 경로통행시간 예측기법들에 비하여 매우 우수하게 나타났다.

주요어 : 첨단자료관리시스템, 대용량 자료, k -최근린 비모수회귀, j -클러스터링, 장기예측, 도로망 소통상태, 도시간 경로통행시간

학 번 : 2008-30675

목 차

I. 서론	1
1. 연구의 배경 및 목적	1
2. 연구의 범위 및 방법	4
3. 연구의 수행체계	5
II. 이론 및 선행연구 고찰	8
1. 교통류 특성과 예측모형의 관계	8
1) 교통류 상태의 진화 특성	8
2) 거시적 교통변수 특성	9
3) 교통류 상태 이론과 예측문제 접근법	11
2. 교통변수 예측에 관한 연구	13
1) 예측의 시·공간 정의	13
2) 입력벡터의 차원	14
3) 예측문제의 정의	15
4) 예측모형의 분류	16
5) 정적(Naive) 기법	17
6) 모수(Parametric) 모형	18
7) 비모수(Non-parametric) 모형	21
8) 기존 교통상태 예측 연구 고찰의 소결	26
3. 구간기반 통행시간 예측모형	29
1) 구간기반 통행시간 예측의 시·공간적 구조	29
2) 구간통행시간 예측 연구	30
3) 구간통행시간 예측모형의 문제점	32
4) 구간기반 경로통행시간 예측모형	34
5) 구간기반 경로통행시간 예측모형의 문제점	37

4. 경로기반 통행시간 예측모형	38
1) 경로기반 통행시간 예측모형의 정의	38
2) 경로기반 통행시간 예측 연구	39
3) 경로기반 통행시간 예측의 문제점	40
5. 기존 통행시간 모형의 한계 및 문제점	41
1) 도착시각 기준 통행시간 모형	41
2) 구간기반 경로통행시간 모형	42
3) 경험기반 경로통행시간 모형	43
4) 인위적 모델링의 한계	45
5) 적용 및 유지·보수의 한계	45
6. 본 연구의 차별성과 기여도	46
1) 연구의 차별성	46
2) 연구의 기여도	48
 III. 대용량 자료기반 통행시간 예측모형 개발	51
1. 모형의 개발과정 정립	51
1) 개발방향의 설정	51
2) 요구조건 분석	52
3) 개발모형의 개념 및 접근법	55
4) 모형의 개발 과정 정립	56
5) 개발모형의 구조	58
2. 데이터 탐색기반의 예측 연구	61
1) KNN-NPR 모형의 구조	61
2) KNN-NPR의 이론적 배경과 적용	62
3) ITS 예측분야의 KNN-NPR 적용 연구	63
3. 입력변수의 거시적 특성분석	65
1) 혼잡의 시·공간적 진화 행태	65
2) 프로브 통행량의 시·공간적 변화	67

3) 프로브 통행량과 속도의 관계	69
4) 유·출입 통행량과 경로통행시간의 관계	72
4. 개발모형의 예측문제 정의	74
1) 현행상태 입력변수의 선정	74
2) 입·출력 변수의 정의	75
3) 예측 문제의 정의	76
5. KN단계 모형의 개발	78
1) 상태벡터의 정의	78
2) 상태간 거리 Metric	83
3) KNN의 DB구조	86
4) KNN 군집구축 알고리즘	87
6. JC단계 모형의 개발	88
1) 입·출력 자료의 정의	88
2) 기존 군집화 모형의 요구수준 적합성 분석	89
3) j-Clustering 모형개발	90
4) 적정 군집수 결정 지표	95
5) JC단계의 출력값	98
6) J-군집화 알고리즘	98
7. FDM단계 모형의 개발	100
1) 의사결정 방법론 정립	100
2) 네트워크 소통상태 예측	102
3) 경로통행시간 예측	103
4) FDM 알고리즘	106
8. 통합 KJC 예측기의 수행구조	107
 IV. 자료구축 방법론	 111
1. 개요	111
1) 배경 및 목적	111

2) 자료구축 방법론의 구성	112
2. 수집대상자료의 특성분석	113
1) 구간통행시간 자료의 특성분석	113
2) 경로통행시간 자료의 특성분석	114
3) 통행시간 추정 방법론의 요구조건 분석	115
3. 기존 방법론 고찰	116
1) 기존 통행시간 추정 방법론 고찰	116
2) 기존 방법론의 문제점 분석	120
3) 해결방안 도출	122
4. 통행시간 추정 방법론 개발	126
1) 집계시간 길이와 최소 표본수의 결정	126
2) 구간통행시간 추정 방법론 개발	128
3) 경로통행시간 추정 방법론 개발	140
5. 자료 구축	152
1) 수집자료 선정	152
2) 수집자료의 내용 및 범위	153
3) 자료구축을 위한 설계	153
4) 자료구축의 과정	157
5) 유·출입 통행량 구축자료의 특성	159
6) 구간통행시간 구축자료의 특성	160
7) 구간통행속도 구축자료의 특성	161
8) 구간 프로브 통행량 구축자료의 특성	162
9) 경로통행시간 구축자료의 특성	163
V. 개발모형의 평가	164
1. 평가 개요	164
2. 평가 설계	165
1) 비교모형의 선정	165

2) 평가지표의 설정	166
3) 경로통행시간 형태의 선정	168
4) 평가를 위한 단 · 중 · 장거리 통행의 설정	170
3. 파라미터 최적화	170
1) 파라미터 최적화 방법	170
2) 파라미터 최적화 결과	172
4. 평가 결과	175
1) 종합평가 결과	175
2) 적용 결과: 비혼잡	180
3) 적용 결과: 오전 \cap 형	187
4) 적용 결과: 오후 \cap 형I	193
5) 적용 결과: 오후 \cap 형II	199
6) 적용 결과: 오전 \wedge 형	205
7) 적용 결과: 오후 \wedge 형	211
8) 적용 결과: 2봉	217
9) 적용 결과: 3봉	223
10) 적용 결과: 주간 지속형	229
11) 적용 결과: 극심한 정체	235
5. 소결	241
1) 교통정보 예측에 관한 소결	241
2) 모형개발에 관한 소결	243
VI. 결론 및 향후 연구	245
1. 결론	245
2. 향후 연구	247

표 목차

<표 2-1> 구간통행시간 예측연구의 고찰 결과	31
<표 3-1> 적정 군집수 결정 모형	96
<표 3-2> JC단계의 출력값	98
<표 3-3> FDM단계 의사결정권의 부여결과의 예	101
<표 4-1> 구간통행속도 그룹군 분류	123
<표 4-2> 도로구간 및 존 설정	154
<표 4-3> 선정된 존간 거리(km)	156
<표 5-1> 단·중·장거리 통행 설정 결과	170
<표 5-2> 장거리 경로통행시간의 종합평가 결과	176
<표 5-3> 중거리 경로통행시간의 종합평가 결과	177
<표 5-4> 단거리 경로통행시간의 종합평가 결과	178

그림 목차

<그림 1-1> 연구의 수행체계	7
<그림 2-1> 교통류 상태의 시-공간적 진화	9
<그림 2-2> 거시적 교통량-밀도 관계	10
<그림 2-3> 시계열 상태의 구조 및 정의	14
<그림 2-4> 예측모형의 분류	16
<그림 2-5> 구간기반 통행시간의 시·공간 예측 영역	29
<그림 2-6> 구간기반 경로통행시간 예측모형의 시·공간	36
<그림 2-7> 경로기반 통행시간의 시·공간 예측 영역	38
<그림 2-8> 출발시각과 도착시각 기준 경로통행시간의 비교	41
<그림 2-9> 출발시각 기준과 동시 경로통행시간의 비교	42
<그림 2-10> 출발시각 기준과 경험 경로통행시간의 비교	43
<그림 2-11> 혼잡의 시간적 주기성과 변동성	44
<그림 3-1> 개발모형의 파라미터값에 따른 오차공간 목표	54
<그림 3-2> 혼잡의 동적 생애주기 및 다양성	55
<그림 3-3> 모형의 개발 과정	58
<그림 3-4> 개발모형(KJC Forecaster)의 구조	60
<그림 3-5> 모수기반 모형과 NPR 모형의 구조	61
<그림 3-6> 혼잡의 시·공간 진화와 패턴	67
<그림 3-7> 구간 프로브 통행량의 시·공간 변화	69
<그림 3-8> 구간 프로브 통행량과 구간통행속도의 거시적 관계	71
<그림 3-9> 구간 프로브 통행량과 구간통행속도의 미시적 관계	71
<그림 3-10> 유·출입 통행량과 경로통행시간의 거시적 관계	73
<그림 3-11> 유·출입 통행량과 경로통행시간의 미시적 관계	73
<그림 3-12> 개발모형의 입·출력 자료구조	75

<그림 3-13> 개발모형의 시·공간 정의	76
<그림 3-14> 상태벡터와 KNN의 개념	78
<그림 3-15> 시간처짐을 고려한 상태벡터의 시·공간 정의	80
<그림 3-16> KNN 군집구축 알고리즘의 DB구조	86
<그림 3-17> KNN 군집구축 알고리즘	87
<그림 3-18> JC단계 모형의 입력값	88
<그림 3-19> 이상치 검지 및 제거의 예	91
<그림 3-20> 절대 쌍 선정결과의 예	92
<그림 3-21> 구축된 하부 군집의 예	94
<그림 3-22> 군집 병합 과정의 예	94
<그림 3-23> 적정 군집수 결정의 원칙	96
<그림 3-24> 적정 군집수 결정의 예	97
<그림 3-25> 최종 군집화 결과의 예	97
<그림 3-26> J-군집화 알고리즘	99
<그림 3-27> IPTT와 EPTT의 예측영역과 경로통행시간 궤적	104
<그림 3-28> DLPTT와 PPTT의 예측영역과 경로통행시간 궤적	105
<그림 3-29> FDM 알고리즘	107
<그림 3-30> 통합 KJC 예측기의 수행구조	110
<그림 4-1> 자료구축 방법론의 구성	112
<그림 4-2> 개별 구간통행시간 자료의 특성	113
<그림 4-3> 개별 경로통행시간 자료의 특성: 중거리 통행	114
<그림 4-4> 개별 경로통행시간 자료의 특성: 장거리 통행	115
<그림 4-5> 개별 구간통행시간의 분포	122
<그림 4-6> 개별 경로통행시간의 분포	124
<그림 4-7> 장거리 경로통행시간의 모호한 경계 특성	124
<그림 4-8> 경유지기반 릴레이 경로통행시간 산출의 개념	125
<그림 4-9> 통행시간 추정오차의 종류	126

<그림 4-10> 이전 시간대 방향 집계시간 확장	129
<그림 4-11> 시간대별 구간통행시간 프로브 변동	130
<그림 4-12> 대표 그룹 중심 탐색법	130
<그림 4-13> 대표 그룹 중심 탐색법의 반복회수-밀도변화 관계	134
<그림 4-14> 구간통행시간 추정 알고리즘	135
<그림 4-15> 대표 그룹 선정 전/후: 구간통행시간	136
<그림 4-16> 제거 대상에 따른 밀도의 변화: 구간통행시간	136
<그림 4-17> 반복회수에 따른 대푯값의 수렴: 구간통행시간	137
<그림 4-18> 구간통행시간 추정결과	138
<그림 4-19> 추정 구간통행시간의 상태 변동량	139
<그림 4-20> 통행거리 vs. 시간대별 경로통행시간 프로브 대수 변동	140
<그림 4-21> 통행거리 vs. 경로통행시간 프로브의 관계	141
<그림 4-22> 전·후방향 집계시간 확장기반 최소 표본수 확보	142
<그림 4-23> 대표 그룹 선정 전/후: 경로통행시간	143
<그림 4-24> 제거 대상에 따른 밀도의 변화: 경로통행시간	144
<그림 4-25> 반복회수에 따른 대푯값의 수렴: 경로통행시간	144
<그림 4-26> 직접 경로통행시간 추정결과: 천안IC→서울TG	145
<그림 4-27> 직접 경로통행시간의 상태 변동량: 천안IC→서울TG	145
<그림 4-28> 직접 경로통행시간 추정결과: 대전IC→천안IC	146
<그림 4-29> 직접 경로통행시간의 상태 변동량: 대전IC→천안IC	146
<그림 4-30> 경유지기반 경로통행시간 산출	147
<그림 4-31> 직접 경로통행시간과 경유지기반 경로통행시간의 비교	149
<그림 4-32> 직접 경로통행시간과 경유지기반 경로통행시간의 검정 결과	149
<그림 4-33> 경유지기반 경로통행시간 추정결과: 대전IC→서울TG	150
<그림 4-34> 경유지기반 경로통행시간의 상태 변동량: 대전IC→서울TG	150
<그림 4-35> 경로통행시간 추정 알고리즘	151
<그림 4-36> 설정된 도로구간의 세부정보	155
<그림 4-37> 도로구간 설정	156

<그림 4-38> 세부 자료별 구축과정	158
<그림 4-39> TCS 통행량 구축자료의 특성	159
<그림 4-40> 구간통행시간 구축자료의 특성	160
<그림 4-41> 구간통행속도 구축자료의 특성	161
<그림 4-42> 구간 프로브 통행량 구축자료의 특성	162
<그림 4-43> 경로통행시간 구축자료의 특성	163
<그림 5-1> 개발모형의 평가 과정	164
<그림 5-2> 변동분석 그래프	168
<그림 5-3> 단조로운 경로통행시간의 형태	169
<그림 5-4> 선별된 경로통행시간의 형태	169
<그림 5-5> 파라미터 최적화 결과: 시나리오 1	173
<그림 5-6> 파라미터 최적화 결과: 시나리오 2	173
<그림 5-7> 최적 파라미터 값의 선정	174
<그림 5-8> 통행거리에 따른 예측모형별 성능 비교 결과	179
<그림 5-9> 소통상태 예측결과: 비혼잡	181
<그림 5-10> 장거리 경로통행시간 예측결과: 비혼잡	182
<그림 5-11> 중거리 경로통행시간 예측결과: 비혼잡	183
<그림 5-12> 단거리 경로통행시간 예측결과: 비혼잡	184
<그림 5-13> 장거리 경로통행시간 예측 변동분석: 비혼잡	185
<그림 5-14> 중거리 경로통행시간 예측 변동분석: 비혼잡	186
<그림 5-15> 단거리 경로통행시간 예측 변동분석: 비혼잡	186
<그림 5-16> 소통상태 예측결과: 오전 \cap 형	187
<그림 5-17> 장거리 경로통행시간 예측결과: 오전 \cap 형	188
<그림 5-18> 중거리 경로통행시간 예측결과: 오전 \cap 형	189
<그림 5-19> 단거리 경로통행시간 예측결과: 오전 \cap 형	190
<그림 5-20> 장거리 경로통행시간 예측 변동분석: 오전 \cap 형	191
<그림 5-21> 중거리 경로통행시간 예측 변동분석: 오전 \cap 형	192
<그림 5-22> 단거리 경로통행시간 예측 변동분석: 오전 \cap 형	192

<그림 5-23> 소통상태 예측결과: 오후 \cap 형I	193
<그림 5-24> 장거리 경로통행시간 예측결과: 오후 \cap 형I	194
<그림 5-25> 중거리 경로통행시간 예측결과: 오후 \cap 형I	195
<그림 5-26> 단거리 경로통행시간 예측결과: 오후 \cap 형I	196
<그림 5-27> 장거리 경로통행시간 예측 변동분석: 오후 \cap 형I	197
<그림 5-28> 중거리 경로통행시간 예측 변동분석: 오후 \cap 형I	198
<그림 5-29> 단거리 경로통행시간 예측 변동분석: 오후 \cap 형I	198
<그림 5-30> 소통상태 예측결과: 오후 \cap 형II	199
<그림 5-31> 장거리 경로통행시간 예측결과: 오후 \cap 형II	200
<그림 5-32> 중거리 경로통행시간 예측결과: 오후 \cap 형II	201
<그림 5-33> 단거리 경로통행시간 예측결과: 오후 \cap 형II	202
<그림 5-34> 장거리 경로통행시간 예측 변동분석: 오후 \cap 형II	203
<그림 5-35> 중거리 경로통행시간 예측 변동분석: 오후 \cap 형II	204
<그림 5-36> 단거리 경로통행시간 예측 변동분석: 오후 \cap 형II	204
<그림 5-37> 소통상태 예측결과: 오전 \wedge 형	205
<그림 5-38> 장거리 경로통행시간 예측결과: 오전 \wedge 형	206
<그림 5-39> 중거리 경로통행시간 예측결과: 오전 \wedge 형	207
<그림 5-40> 단거리 경로통행시간 예측결과: 오전 \wedge 형	208
<그림 5-41> 장거리 경로통행시간 예측 변동분석: 오전 \wedge 형	209
<그림 5-42> 중거리 경로통행시간 예측 변동분석: 오전 \wedge 형	210
<그림 5-43> 단거리 경로통행시간 예측 변동분석: 오전 \wedge 형	210
<그림 5-44> 소통상태 예측결과: 오후 \wedge 형	211
<그림 5-45> 장거리 경로통행시간 예측결과: 오후 \wedge 형	212
<그림 5-46> 중거리 경로통행시간 예측결과: 오후 \wedge 형	213
<그림 5-47> 단거리 경로통행시간 예측결과: 오후 \wedge 형	214
<그림 5-48> 장거리 경로통행시간 예측 변동분석: 오후 \wedge 형	215
<그림 5-49> 중거리 경로통행시간 예측 변동분석: 오후 \wedge 형	216
<그림 5-50> 단거리 경로통행시간 예측 변동분석: 오후 \wedge 형	216

<그림 5-51> 소통상태 예측결과: 2봉	217
<그림 5-52> 장거리 경로통행시간 예측결과: 2봉	218
<그림 5-53> 중거리 경로통행시간 예측결과: 2봉	219
<그림 5-54> 단거리 경로통행시간 예측결과: 2봉	220
<그림 5-55> 장거리 경로통행시간 예측 변동분석: 2봉	221
<그림 5-56> 중거리 경로통행시간 예측 변동분석: 2봉	222
<그림 5-57> 단거리 경로통행시간 예측 변동분석: 2봉	222
<그림 5-58> 소통상태 예측결과: 3봉	223
<그림 5-59> 장거리 경로통행시간 예측결과: 3봉	224
<그림 5-60> 중거리 경로통행시간 예측결과: 3봉	225
<그림 5-61> 단거리 경로통행시간 예측결과: 3봉	226
<그림 5-62> 장거리 경로통행시간 예측 변동분석: 3봉	227
<그림 5-63> 중거리 경로통행시간 예측 변동분석: 3봉	228
<그림 5-64> 단거리 경로통행시간 예측 변동분석: 3봉	228
<그림 5-65> 소통상태 예측결과: 주간 지속형	229
<그림 5-66> 장거리 경로통행시간 예측결과: 주간 지속형	230
<그림 5-67> 중거리 경로통행시간 예측결과: 주간 지속형	231
<그림 5-68> 단거리 경로통행시간 예측결과: 주간 지속형	232
<그림 5-69> 장거리 경로통행시간 예측 변동분석: 주간 지속형	233
<그림 5-70> 중거리 경로통행시간 예측 변동분석: 주간 지속형	234
<그림 5-71> 단거리 경로통행시간 예측 변동분석: 주간 지속형	234
<그림 5-72> 소통상태 예측결과: 극심한 정체	235
<그림 5-73> 장거리 경로통행시간 예측결과: 극심한 정체	236
<그림 5-74> 중거리 경로통행시간 예측결과: 극심한 정체	237
<그림 5-75> 단거리 경로통행시간 예측결과: 극심한 정체	238
<그림 5-76> 장거리 경로통행시간 예측 변동분석: 극심한 정체	239
<그림 5-77> 중거리 경로통행시간 예측 변동분석: 극심한 정체	240
<그림 5-78> 단거리 경로통행시간 예측 변동분석: 극심한 정체	240

I. 서론

1. 연구의 배경 및 목적

우리나라는 국토의 균형발전을 위해 국가의 주요 기반시설이 지방으로 분산되고 있으며, 삶의 수준이 향상되면서 여가통행은 꾸준히 증가하고 있다. 도시교통의 행태적 관점에서 이는 향후 도시간 중·장거리 업무 및 여가 통행이 지속적으로 증가할 것을 의미한다. 그리고 중·장거리 자동차 통행의 경우, 대다수의 운전자는 네비게이션에 의해 제공되는 경로 및 통행시간 정보에 대한 의존성이 매우 높다. 따라서 정확한 중·장거리 통행시간 정보의 예측과 제공은 21세기 정보화시대에 매우 주요한 정보요소라 할 수 있으며, 이는 지능형교통체계(Intelligent Transport System, 이하 ITS)를 통해 달성된다. 일련의 ITS 정보체계(정보수집→정보가공→장래 상태 예측→교통제어와 교통정보제공)에서 교통상태 예측시스템은 ITS 하위 시스템으로서 매우 주요한 역할을 수행한다. 특히, 교통상태 예측시스템은 첨단교통정보시스템(Advanced Traffic Information System, 이하 ATIS)의 핵심 하위시스템이다.

20세기 말 국내에 ITS가 도입된 이후로 통행시간 예측 기술은 보다 정교해지고 고도화되었으나, 장래 5~15분 이내의 교통상태를 예측하는 단기예측(Short-term forecasting) 기법에 대부분을 의존하고 있다. 따라서 단기예측기법은 통행시간의 길이가 30분 이상인 중·장거리 통행시간 정보의 예측에 있어 시·공간적 불일치의 문제가 발생한다. 이러한 기술적 한계로 인해 네비게이션을 이용하는 다수의 운전자는 네비게이션으로 제공되는 출발지에서 목적지까지의 경로정보는 신뢰하지만 소요시간정보는 신뢰하지 않고 있는 실정이다.

21세기에 들어 ATIS 분야는 첨단 IT기술을 기반으로 다양한 정보수집 매체를 통하여 다양한 정보제공 서비스를 일반시민에게 제공하고 있다. 이러한 다양한 교통정보서비스의 이면에는 안타까운 현실이 있다. 21세기 IT기술을 기반으로 하는 시스템의 Hardware와 정보제공 콘텐츠의

기술적 진보에도 불구하고, 제공서비스의 핵심요소인 장래 교통상태는 이러한 기술적 발전과 병행하지 못하고 있는 실정이다. 일례로 소수 민간 교통정보서비스 업체와 한국도로공사를 제외하면 대부분 실시간 정보에 의존하고 있는 실정이다.

한편으로 ITS의 교통정보 예측분야는 20세기 말의 1차 도약기를 거쳐, 현재 성숙기를 지나, 21세기 2차 도약을 달성해야 한다는 의미이기도 하다. 그러나 2차 도약기에 접어들기 위해서는 현재의 예측 기술이 넘지 못하고 있는 (ITS 정보제공 측면에서) 중·장기 장래 상태에 대한 불확실성(Uncertainty)을 극복함으로서 예측의 범위(Forecasting horizon)을 확장해야 한다. 수많은 단기예측모형이 개발 및 고도화되었으나, van Hinsbergen et al.(2007)이 지적한 바와 같이 타 모형에 비하여 매우 우수한 성능을 가진 모형은 없다고 지적하였다. 결국, Chang et al.(2012b)가 지적한 바와 같이 21세기 ITS 예측분야의 개척 영역중 하나는 수용 가능한 오차의 수준이내에서 예측영역의 시·공간적 확장이라 판단된다.

오늘날의 도시교통 정보분야는 대용량 자료로 대변되는 정보 홍수의 시대를 겪고 있으며, 대용량 자료의 수집원은 스마트 단말(Smart device)이라해도 과언은 아니다. 국내의 스마트폰 보급률은 70%을 넘어섰으며, 한국도로공사의 DSRC 단말기 보급률은 800만대를 넘어섰다. 이러한 정보수집체계의 급속한 변화는 기존 ITS 예측영역의 기술적 한계를 극복할 수 있을 뿐만 아니라 새로운 영역을 개척할 수 있는 기회를 제공하고 있다.

따라서 본 연구의 목적은 스마트 단말에서 수집되는 방대한 양의 정보를 수집·가공·관리하는 첨단자료관리시스템(Advanced Data Management System, 이하 ADMS)의 대용량 자료를 이용하여 장래 6시간까지 도로망 소통상태를 네트워크 기반으로 예측함과 동시에 예측된 도로망 소통정보를 이용하여 단·중·장거리 도시간 경로통행시간 정보를 동적으로 산출하는 예측기(Forecaster)를 개발하는데 있다. 그리고 실제 대용량 자료환경에서 개발된 예측기의 성능과 적용 가능성을 종합적으로 평가하도록 한다. 그리고 본 연구의 제목과 관련된 연구의 내용은 다음과 같다.

“대용량 자료를 이용한”은 ADMS와 같은 자료관리시스템에 탑재되어 시스템의 방대한 데이터베이스 안에 내재된 다양한 형태의 장래상태 정보를 지식탐색(Knowledge Discovery, 이하 KD)의 과정을 통해 추출한 후, 추출된 지식을 이용한 장래 교통상태 예측을 의미한다. 따라서 본 연구의 예측모형은 KD 접근법을 기반으로 개발되며, 시스템의 탑재와 운영을 고려하여 모형의 구조변경, 입력자료와 출력자료의 용이한 추가 및 삭제, 파라미터 값의 정산, 모형의 연산수행속도, 데이터베이스에 포함된 결측자료의 처리 등 자료관리시스템의 하위 시스템으로 갖추어야할 요구조건을 충족하도록 개발된다.

“네트워크 기반”은 단일 도로 구간¹⁾이나 일련의 연속된 몇 개의 도로 구간이 아닌 도로망을 구성하는 전체 도로구간에 대한 소통상태의 예측을 의미한다. 다시말해, 기존의 구간기반 통행시간 예측의 개념에서 벗어나 도로망의 전체 도로구간 소통상태를 개발모형의 입력값으로 이용하여 도로망의 전체 도로구간 소통상태를 예측함을 의미한다.

“도시간 경로통행시간 예측”은 기존 단기 통행시간 예측모형의 시간적 예측영역(보통 15분 이내의 장래)의 한계를 극복하고 장래 6시간까지 예측영역을 확장함을 의미한다. 이는 도시간 통행의 많은 부분은 30분 이상의 중·장거리 통행이기 때문이다. 단기 통행시간 예측모형의 경우, 단일 구간에 대한 소량의 시계열자료를 이용하여 장래상태를 예측하였기 때문에 장래 상태에 대한 불확실성을 극복하지 못 하였다. 따라서 수용 가능한 예측오차의 범위내에서 30분 이상의 중·장거리 경로통행시간을 예측할 수 없는 근본적 한계가 있었다. 이러한 문제를 극복하기 위하여 본 연구에서는 도로망 차원에서 장래 소통상태 설명이 가능하도록 네트워크 차원의 방대한 시스템 입력자료를 이용하여 장래의 불확실성을 최소화함으로써 예측영역을 장래 6시간까지 확장하고, 중·장거리 통행시간의 정보생성이 가능하도록 하였다.

1) 본 연구에서 (도로)구간은 도로구간(Road section), 도로망을 구성하는 링크(Link), 그리고 경로(Path)를 구성하는 링크를 말하며, 이후 (도로)구간으로 통일하도록 한다.

2. 연구의 범위 및 방법

본 연구에서 개발되는 예측기는 ADMS와 같은 자료관리시스템에 탑재되어 방대한 양의 현행 및 이력 자료를 이용하여 도로망의 소통상태를 예측하고, 예측된 소통상태를 이용하여 도시간 경로통행시간을 예측한다. 그리고 예측기의 개발을 위한 방법론의 개발에 관한 연구의 범위 및 내용은 다음과 같다.

첫째, 개발 방법론은 방대한 양의 시스템 자료를 이용하도록 개발된다. 시스템 자료는 ADMS와 같은 자료관리시스템에 운영/관리되고 있는 자료를 의미한다.

둘째, 통행시간 예측모형의 예측영역(Forecasting horizon)은 현재 시점에서 5분 간격으로 72개 시간대까지 장래 즉, 장래 6시간까지 도로구간별 통행시간을 예측하고, 예측된 구간통행시간 자료를 이용하여 출발 시각 기준의 단·중·장거리의 경로통행시간을 예측하도록 한다.

셋째, 본 연구에서 개발되는 대용량 데이터 기반의 예측 기(Forecaster)는 3단계 모형으로 구성된다. 1단계는 이력 데이터베이스의 탐색과정을 통해 장래 상태를 설명하기 위한 k 개의 이웃(k -nearest neighbor, 이하 KN)로 구성되는 1차 의사결정 집단을 구성하는 단계로서 일종의 정보추출 모형이 개발·적용된다. 2단계는 1단계에서 구축된 1차 의사결정 집단을 입력자료로 군집화(Clustering) 모형을 이용하여 j 개의 군집으로 구성되는 다수의 2차 의사결정 집단을 결정하는 JC(j -clustering)단계로서 일종의 결합(Merging)기반 군집화 모형이 개발된다. 3단계는 2단계에서 구축된 j 개의 군집 즉, 2차 의사결정 집단들 중 최종 의사결정 집단을 결정하고, 해당 집단을 구성하는 m 개 멤버($m < k$, $m > 1$)의 과거 상태를 이용하여 장래 상태를 예측하는 단계이다. 3단계에서 예측모형은 네트워크 소통상태 예측모형과 경로통행시간 예측모형으로 각각 개발된다.

넷째, 예측 방법론의 평가를 위한 선행단계로 자료관리시스템의 자료 지원 환경을 조성하기 위하여 방대한 양의 이력자료를 구축하도록 한다. 구축된 데이터 환경을 기반으로 자료관리시스템에 탑재/운영되는 데이터 기반의 도로망 소통상태 및 도시간 경로통행시간 예측모형을 개발하도록

한다. 개발된 예측모형은 실제 자료를 이용하여 성능을 평가하도록 한다.

다섯째, 연구의 공간적 범위²⁾는 DSRC 단말 자료가 수집되는 고속도로 노선이다. 그리고 개발 방법론의 단·중·장거리 통행시간 예측 능력을 평가하기 위한 사례대상 구간은 경부고속도로 양재IC - 대전IC 구간(142.3km)이다.

여섯째, 수집자료의 범위는 한국도로공사에서 운영 중인 DSRC 단말에서 수집되는 구간 통행속도와 통행시간, DSRC 구간 프로브 통행량, 그리고 톨게이트 유·출입 통행량이다. 수집된 자료를 이용하여 과거 구간 통행시간과 주요 도시의 나들목(Interchange, 이하 IC) 및 영업소(Tollgate, 이하 TG)간 경로통행시간 데이터베이스를 구축하여 ADMS 데이터 환경을 구축하도록 한다.

마지막으로 구축된 과거 데이터베이스를 이용하여 개발된 3단계 모형의 통합모형인 KJ 예측기의 성능평가를 수행하도록 한다. KJC 예측기의 성능평가는 혼잡의 형태에 따른 다양한 분석 시나리오에 대하여 예측의 정확도, 예측기의 수행속도 등을 통해 종합적으로 수행된다.

3. 연구의 수행체계

본 연구는 이론 및 선행연구 고찰, 대용량 자료기반 통행시간 예측모형 개발, 자료구축 방법론, 개발모형의 적용 및 평가, 결론 및 향후연구로 수행되며, 연구의 수행체계는 <그림 1-1>과 같다.

2장에서 거시적 교통류 변수 예측의 이론적 고찰을 기반으로 구간기반 통행시간 예측모형과 경로기반 통행시간 예측에 관한 연구를 고찰한다. 그리고 선행연구의 한계 및 문제점을 도출한 후, 선행연구의 한계를 극복하기 위한 연구의 방향과 접근법을 설정하고 본 연구의 차별성과 기여도를 도출한다.

3장에서 모형의 개발과정과 개발모형을 설명하도록 한다. 모형의 개발

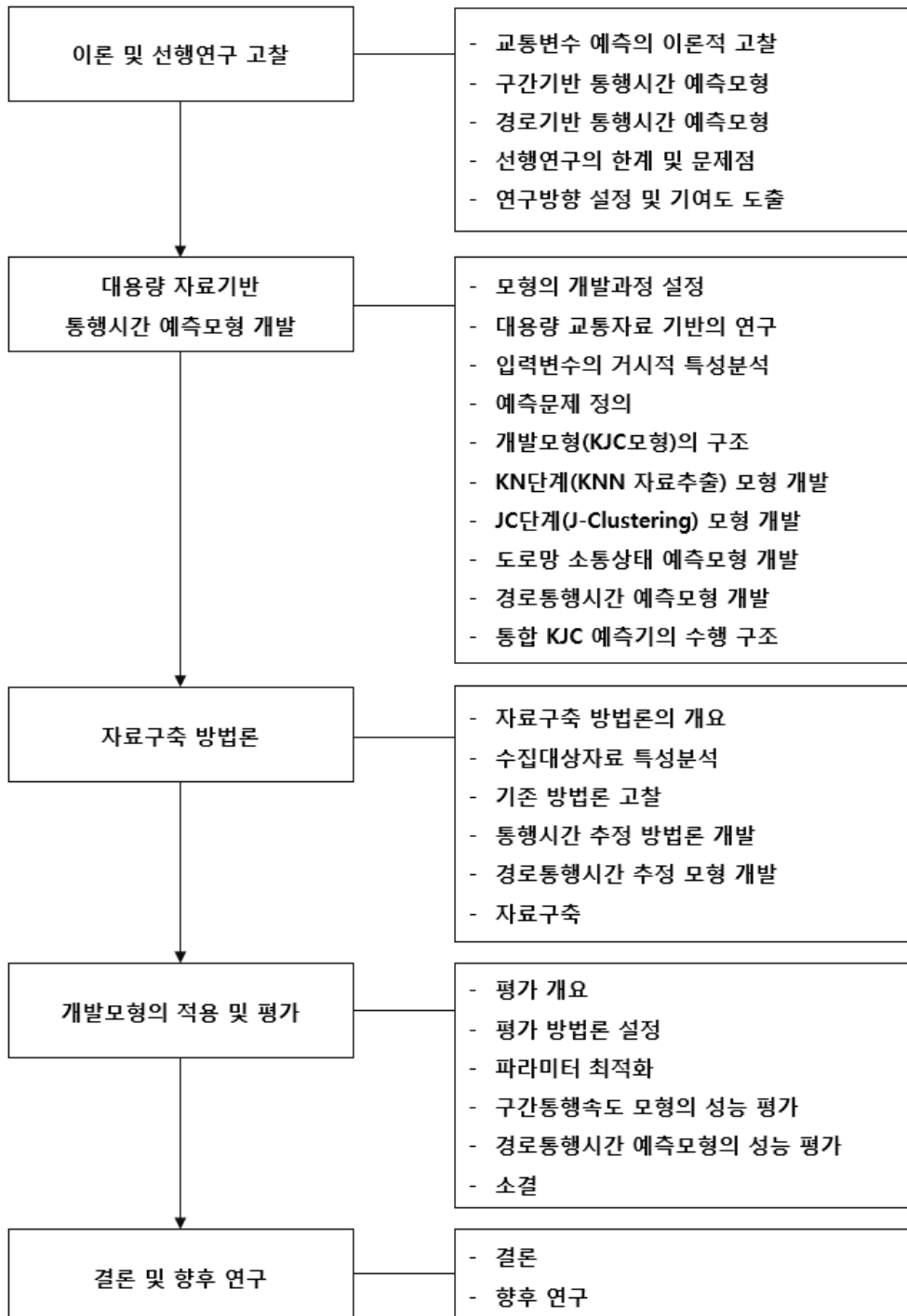
2) 개발모형의 공간적 범위는 단속류와 연속류를 포함하는 도로망이다. 그러나 자료의 수집 및 모형평가를 위한 목적으로 본 연구에서 공간적 범위는 고속도로로 한정한다.

을 위하여 모형의 개발 목표와 개발 과정을 설정하고 대용량 교통자료 기반의 기존연구를 수행한다. 모형의 개발은 입력변수의 거시적 특성분석, 예측 문제의 정의, 개발모형의 구조 정립, KN(k -nearest neighbor, k 개의 최근린 이웃)단계의 자료추출 모형개발, JC(j -Clustering, j 개의 과거 패턴 분류)단계의 패턴인식 모형개발, KJC단계에서 추출-분류된 과거 패턴을 이용하여 예측 의사결정을 수행하는 FDM(Forecasting decision making) 단계로 구성되며, FDM단계는 도로망 소통상태 예측모형과 경로통행시간 예측모형으로 구성된다. 마지막으로 통합 KJC 예측모형의 수행구조와 통합 알고리즘을 설명한다.

4장에서는 개발 방법론의 성능평가를 위한 데이터베이스의 구축 방법론에 대하여 설명하도록 한다. 이는 개발모형의 성능은 전적으로 입력자료의 질과 다양성에 의존하기 때문이다. 입력 데이터베이스 구축은 자료구축 방법론의 개요, 수집대상 자료의 특성분석, 기존 통행시간 추정 방법론의 고찰, 통행시간 추정방법론 개발, 자료구축으로 구성된다. 그리고 통행시간 추정 방법론은 구간 및 경로 통행시간 추정모형으로 각각 개발된다. 마지막으로 구축 자료별 자료구축 방법론과 더불어 구축자료의 특성을 예측의 관점에서 설명하도록 한다.

5장에서는 구축된 데이터베이스를 이용하여 개발모형의 성능을 종합적으로 평가하도록 한다. 개발모형의 평가는 적용 및 평가의 목표를 수립한 후 평가 방법론을 설정하도록 한다. 평가 방법론은 비교모형의 선정, 평가지표의 설정, 다양한 경로통행시간 형태의 선정 등으로 구성된다. 파라미터 최적화 과정을 통해 최적 입력변수와 파라미터 값을 결정하도록 한다. 평가결과는 거시적 종합평가와 혼잡의 형태에 따른 미시적 개별평가로 수행된다. 그리고 성능평가 결과를 통해 도출된 분석 결과(Findings)와 시사점을 소결을 통해 기술하도록 한다.

마지막으로 6장에서 본 연구의 결론 및 시사점, 개발된 예측기의 활용 방안 및 한계, 그리고 개발된 예측기의 고도화를 위한 향후 연구를 제시하도록 한다.



<그림 1-1> 연구의 수행체계

Ⅱ. 이론 및 선행연구 고찰

1. 교통류 특성과 예측모형의 관계

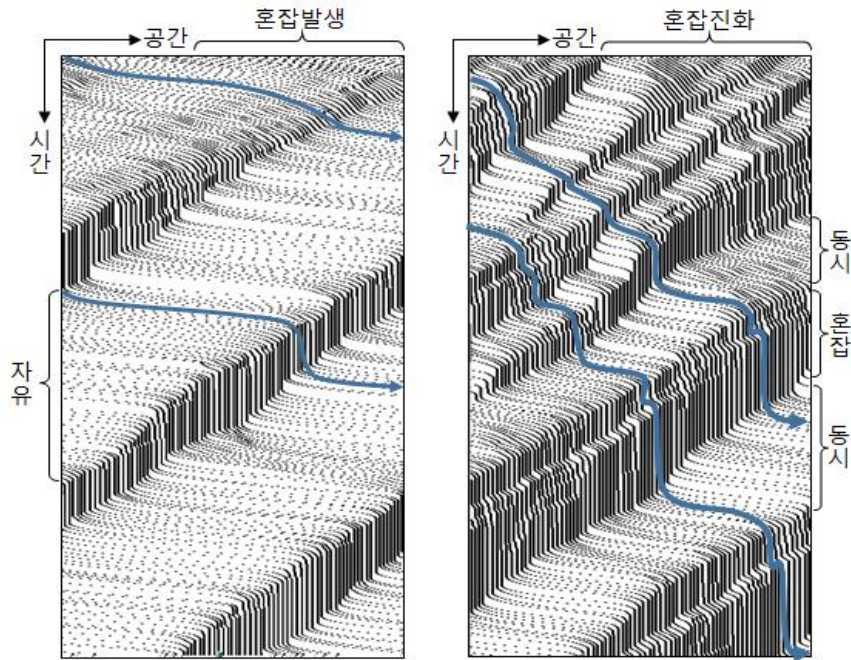
1) 교통류 상태의 진화 특성

일반적으로 예측과정은 대상 시스템의 시·공간적 상태변화에 대한 깊은 이해에 기초한 인위적 모델링 과정을 통해 장래의 불확실성(Uncertainty)을 극복하기 위한 일련의 과정이라 정의할 수 있다. 그리고 특정 시스템의 상태를 예측하기 위해서는 상태의 시·공간적 진화(Evolution)에 대한 이해가 선행되어야 한다. 필요하다. 그러므로 기존 이론과 선행연구의 고찰에 앞서, 거시적 교통류 시스템의 상태이론에 대한 고찰을 선행하도록 한다. 거시적 교통류 상태의 특성에 초점을 둔 이유는 일반적으로 교통상태 예측은 거시적 변수를 대상으로 하기 때문이다.

교통류 시스템의 시·공간적 상태 진화는 현시(Phase) 교통류(Kerner; 1996, 2002)로 설명할 수 있다. 현시교통류에서 교통류는 자유교통류(F), 동시교통류(S), 혼잡교통류(J)의 3개 상태로 구분되며, 교통류의 상태진화 형태 3개 현시의 조합으로 설명된다. 예를 들어 현시의 전환은 $[F \rightarrow J \rightarrow F]$, $[F \rightarrow S \rightarrow J \rightarrow F]$, $[F \rightarrow S \rightarrow J \rightarrow S \rightarrow J \rightarrow S \rightarrow F]$ 등으로 전환하면서 긴 이동 혼잡(Wide-moving jam, 이하 WMJ)으로 발전한다. <그림 2-1>은 혼잡의 발생에서 WMJ로 진화해가는 과정을 차량의 시·공간 궤적(Trajectory)으로 보여주고 있다. 용량을 넘어선 초기 상태의 차량의 수요는 국부적 동시교통류 상태에서 지체를 발생시킨다. 발생한 지체는 혼잡으로 급속히 성장해가면서 F-J단계인 자유교통류와 혼잡교통류로 양분화되고 있다. 이후 혼잡이 성장함에 따라 S-J단계인 동시교통류와 혼잡교통류로 발전하면서 WMJ를 형성한다.

WMJ는 매우 동적이며 다양한 형태를 보이는 것으로 알려져 있다. 따라서 15분 이내의 단기예측을 극복하고 30분~6시간 정도의 중·장기 교통상태 예측을 수행하기 위해서는 단일 도로구간이 아닌 도로망에서 교

통수요와 속도의 변화 즉, 거시적 소통상태를 고려해야 할 것으로 판단된다.



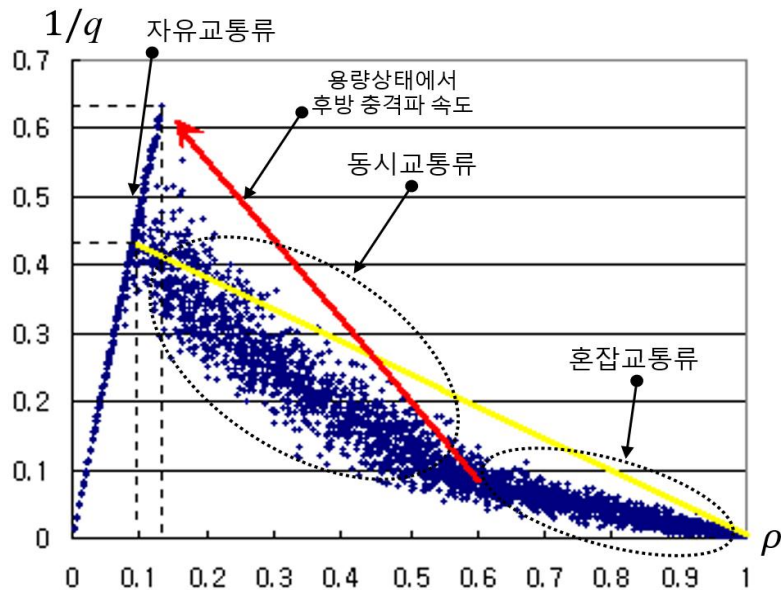
<그림 2-1> 교통류 상태의 시-공간적 진화

2) 거시적 교통변수 특성

교통류 시스템의 상태변화는 거시적 교통류 변수인 교통량(q , 대/단위시간), 속도(v , 단위거리/단위시간), 밀도(ρ , 대/단위거리)로 설명된다. <그림 2-2>는 역 λ -형태의 $q-\rho$ 의 관계를 보여주고 있으며, 자유교통류 상태에서 $q-\rho$ 의 관계는 매우 동질적(Homogeneous)한 상태이다. 그리고 동시(Synchronized) 교통류는 비 동질적(Non-homogeneous) 상태를 보이며, 혼잡 교통류는 다시 동질적 상태를 보인다(장현호 등, 2004; Chang et al., 2005, 2007; Shah et al., 2008). 따라서 연속류의 경우 2개의 불안정 상태(Meta-stable state)가 존재하며, 3개 현시 교통류인 자유교통류, 동

시교통류 그리고 혼잡교통류로 구분된다.

교통류 상태의 시간적 진화 즉, 시계열 자료의 변동 측면에서 정적상태(Stationary condition)와 비정적상태(Non-stationary condition)로 살펴보면 다음과 같다. 정적상태는 $F \rightarrow F$, $(S \rightarrow S)$, 그리고 $J \rightarrow J$ 으로 상태가 유지되는 경우로서 통행속도 예측의 관점에서 볼 때 높은 예측 정확도를 보인다. 반면 교통류 상태가 $F \rightarrow S$, $F \rightarrow J$ 그리고 $S \rightarrow J$ 으로 경계상태(Boundary condition)가 변화하는 경우, 예측모형이 사전에 경계상태를 설명하지 못하면 예측오차는 급격히 증가한다. 따라서 고속도로의 가다 서다 현상과 교통신호가 운영되는 단속류의 경우, 반복적으로 경계상태가 변화하기 때문에 예측 오차는 증폭된다. 또한 교통변수의 경우, 예측의 정확도는 속도, 교통량, 점유율의 순으로 낮게 나타나지만 통행속도의 경우 속도 감소에 따라 통행시간이 급격히 증가함으로 예측오차는 높게 나타난다.



<그림 2-2> 거시적 교통량-밀도 관계: Chang et al.(2007)

3) 교통류 상태 이론과 예측문제 접근법

교통류의 상태진화(State evolution)는 교통상태 예측에 매우 주요한 이론적 근거를 제시한다. 이는 대다수의 교통류 예측모형이 교통류 상태의 특정 시·공간적 진화행태를 근거로 개발되었기 때문이다. 교통류 상태변화에 대한 명확한 정의는 내려지고 있지 않지만, 3개의 주된 주장은 확률적(Stochastic) 상태, 혼돈(Chaotic) 상태, 그리고 확률적 상태보다 혼돈 상태가 강한 혼재된(Mixed) 상태이다(Chang et al. 2012; Yoon and Chang, 2014).

확률적 상태의 주장은 Box-Jenkins(1976) 기법 기반의 시계열 예측모형인 ARIMA(Autogressive Intergrated Moving Average, 이하 ARIMA) 계열을 이용한 연구에서 제기되고 있다. 이는 상태의 정규성을 가정하는 확률적 상태이론에 기초하여 교통류 상태진화를 시계열 예측문제로 모델링을 하기 위한 주장이다. Smith et al.(2002)의 연구는 계절 ARIMA(Seasonal ARIMA, 이하 SARIMA)의 단일 시간대 교통량 예측 정확도가 k -최근린 이웃 비모수 회귀(k -nearest neighbor non-parametric regression, 이하 KNN-NPR)의 예측 정확도보다 근소하게 우수한 결과를 보였으며, 이는 교통류 상태의 진화방식이 확률적이라는 간접적 근거라고 지적하였다. 그러나 연구의 결론에서 모형의 수행속도와 적용성을 고려하여 판단하면 KNN-NPR이 우수하다고 지적하였으며, 지속적인 향후 연구의 필요성을 언급하였다. 아이러니 하게도 Smith의 대다수 연구는 대용량 이력자료를 이용하는 KNN-NPR을 기반으로 수행되었다(Smith 1995; Smith and Demetsky, 1995; Smith and Demetsky, 1996; Smith and Demetsky, 1997; Smith and Oswald, 2003).

반면, Vlahogianni et al.(2005)는 시계열 상태의 변동량이 급격히 변화하는 경우 신경망(Neural network)과 같은 비선형 모형이 상태진화의 비선형성을 설명할 수 있기 때문에 장래상태의 예측에 적합하며, 통계기반 시계열 분석기법인 Kalman Filter 기법(1961)과 ARIMA계열 모형은 상태의

예측에 실패하였다고 언급하였다. 이러한 근거로 Vlahogianni et al.(2006)은 교통류 상태 진화의 특성을 정지성(Stationarity), 비정지성(Non-stationarity), 선형성(linearity), 비선형성(Non-linearity), 초기 결정성(Initial deterministicity), 혼돈(Chaos), 현시전환이동(Transitional movements) 상태로 분석하였다. 그리고 교통류 상태의 시계열적 진화상태는 정지성과 선형성 보다 비정지성, 비선형성, 초기 결정성과 혼돈 상태가 매우 강하다고 언급하였다. 따라서 확률적 상태이론을 기반으로 하는 기존의 (통계기반) 모형은 동적 교통류 상태의 장래 상태예측에 있어 불합리한 결과를 초래할 수 있으며, 인공신경망 같은 비선형 모형의 예측력이 우수하다고 주장하였다.

Altman(1992)는 교통류 상태의 진화형태를 혼재된 상태라 언급하였다. 혼재된 상태의 주장은 NPR를 이용한 교통상태의 예측관련 연구에서 주로 제기되고 있으며, 이는 NPR 접근법이 혼돈 상태이론(Chaotic state theory)에 근거하고 있기 때문이다. Chang et al. (2012_{a,b})는 KNN-NPR을 이용한 다중시간대 교통량의 예측 정확도가 SARIMA의 단일 시간대 예측의 정확도보다 우수한 결과를 근거로 교통류 상태의 진화 행태는 혼돈 또는 혼재된 상태라고 주장하였다. 이는 Smith et al.(2002)의 연구와 상반된 결과이다. 또한 Yoon and Chang(2014)는 교통류 상태의 진화행태가 급격한 변동을 보이는 단속류의 교통량 예측에 있어 KNN-NPR의 예측성능이 적합하며, 확률적 상태이론에 근거한 시계열 예측모형인 Kalman Filter와 SARIMA는 연속적인 상태의 변곡점에서 예측 실패(Forecasting failure)를 피할 수 없음을 보였다. 따라서 단속류의 교통상태 진화행태는 혼재 또는 혼합된 상태라고 주장하였다.

이상에서 고찰한 교통상태의 시·공간적 진화 행태는 예측의 관점에서 볼 때 특정 상태이론을 근간으로 예측 접근법의 옹호와 비판의 영역으로 판단된다. 공통된 사실은 교통류의 진화 행태는 시간과 공간적으로 매우 다양하게 변화하는 동적인 혼재 상태이며, 최근의 연구경향은 확률적 상태이론 보다는 혼돈 또는 혼재된 상태이론을 지지하는 경향이 높다고 판단된다.

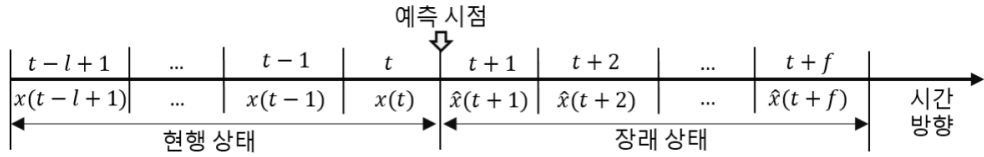
2. 교통변수 예측에 관한 연구

1) 예측의 시·공간 정의

일반적으로 시스템의 시간적 상태는 연속적(Continuous)이지만, 상태의 재구성(Reconstruction)을 이용한 분석에서는 시계열적으로 이산적인 상태(Discrete state)를 이용한다. 이러한 시계열 상태는 예측의 영역에 있어 예측의 시·공간을 정의하는 상태벡터(State vector)로 이용되며, <그림 2-3>은 시각(t)에서 예측의 시간적 정의를 보여주고 있다.

상태의 변화는 예측 시점(Forecasting point, t_f)을 기준으로 시간적인 전/후로 즉, 현행(또는 과거)과 미래로 구분되며, 단위 시간의 길이(Length of time interval, 이하 TL)을 이용하여 이산적 상태로 정의된다. TL 은 예측하고자 하는 교통변수의 용도에 따라 보통 30초~15분 정도로 설정하게 되며, 통행시간 예측의 경우 1~5분을 주로 이용한다. 그리고 각각의 분할된 시간 영역은 시간대(Time interval, 이하 t)으로 정의한다. 시간대가 t_f 보다 이전 시간인 경우 l 개의 시계열 개수인 Embedding size, 즉 lag의 개수를 이용하여 $[(t-l+1), \dots, (t-1), (t)]$ 로 정의하며, 일반적으로 현행 상태벡터의 시간 영역으로 정의한다. 그리고 t_f 보다 미래의 t 는 예측영역(Forecasting horizon, $F=[1, 2, \dots, f-1, f]$, $f>0$)인 f 개의 시간대를 이용하여 $[(t+1), (t+2), \dots, (t+f)]$ 로 정의하며 장래 시간 영역으로 정의한다.

현행 시간영역의 시계열 상태는 주로 개별 현행 시간대의 상태 평균값을 이용하여 (현행) 상태벡터 $X=[x(t), x(t-1), \dots, x(t-l+1)]$ 로 정의되며, 장래 상태벡터는 $\hat{X}=[\hat{x}(t), \hat{x}(t+1), \dots, \hat{x}(t+f)]$ 로 정의된다. 따라서 예측을 수행하기 위해서는 상태벡터를 우선적으로 설정해야 한다. 일반적으로 현행 상태의 시계열자료의 개수 즉, lag의 개수는 최적화 파라미터이다. 그리고 장래 상태의 경우, $f=1$ 인 경우 단일 시간대 예측, $f>1$ 인 경우 다중 시간대 예측으로 정의한다. 추가로 단일 시간대와 다중 시간대 예측을 포함하여 이를 구분하지 않고 단기예측(Short-term prediction)으로 정의하고 있다.



<그림 2-3> 시계열 상태의 구조 및 정의

2) 입력벡터의 차원

입력 상태벡터 또는 현행 상태벡터는 대상 시스템의 상태변화를 재구성(reconstruction)하게 된다. 모형 개발자의 정의에 따라 다양한 상태벡터를 구성할 수 있으며, 상태벡터의 정의는 무한하다(Smith et al. 2002). 그리고 상태벡터의 이산적 상태공간의 길이를 구성하는 시계열 자료의 개수인 Embedding size(l)에 관한 명확한 연구는 제시되고 있지 않고 있다. 대표적인 정의는 Takens(1981)의 정의에 의존하는 경우가 많으며, 주어진 L -차원에서 식(2-1)과 같다.

$$l \geq 2L+1 \quad \text{식(2-1)}$$

Takens의 정의에 의하면 $L=1$ 인 경우 l 은 3 이상에서, $L=2$ 인 경우 l 은 5이상에서 최적의 상태벡터를 구성한다. 다시 말해 $l \geq 2L+1$ 의 조건에서 상태벡터는 시스템의 상태진화 즉, 방향성과 변동량을 최적으로 재구성한다는 의미이다. 반면, Packard et al.(1980)은 식(2-2)의 경우가 존재함을 보였으며, $L=1$ 인 경우 $l=\{1, 2\}$ 에서 상태의 변화를 재구성할 수 있음을 의미한다. 또한 최적의 l 값을 정의하는 일반적인 정의는 존재하지 않는다고 Packard et al.(1980)은 지적하였다. 그러므로 Embedding size는 상태벡터를 이용하는 예측모형의 파라미터 최적화 영역이라고 판단된다.

$$l < 2L+1, l > 0 \quad \text{식(2-2)}$$

일반적으로 l 값은 전역 최적화(Global optimization) 과정을 통하여 추정된다. l 값과 예측오차의 관계는 오목형(Concave)의 관계로서, l 값이 증가함에 따라 예측오차는 급격히 감소하며 오차가 최소화되는 최적 l 값 영역에 수렴한 후 서서히 증가한다(Chang et al. 2010, 2012_{a,b}; Yoon and Chang, 2014). 이는 상태벡터가 재구성하는 현행상태와 유사한 과거 패턴의 경계가 명확하건 명확하지 않건 패턴이 존재함을 의미한다. 그리고 Chang et al.(2012_a)는 단일 구간 또는 단일 지점의 교통상태 예측에 있어 예측 시간대의 길이를 확장함에 따라 l 값의 영역은 확장되고 예측 오차는 증가하며, 예측 시간대의 길이를 지속적으로 확장하면 l 값은 더 이상 증가하지 않음을 보였다. 이는 l 값을 지속적으로 증가시킨다 해도 장래 상태의 불확실성을 지속적으로 감소시킬 수 없음을 의미한다. 이러한 상태벡터의 크기를 결정하는 l 값과 불확실성의 관계는 단기 예측기법의 예측영역 확장 문제와 밀접하게 관련된다.

3) 예측문제의 정의

예측문제는 인류 문명의 발달과 함께 지속적으로 해결하고자 하는 문제중 하나라 할 수 있다. 단일 시스템의 상태변화 예측 문제(Forecasting problem)는 현 상태의 정보를 이용하여 장래 상태의 불확실성을 극복 또는 최소화하기 위한 일련의 과정이며, 식(2-3)과 정의될 수 있다. 예측 시점(t)에서 주어진 (현행) 상태벡터(X_c)를 이용하여 장래 시간대 $t+f$ 의 상태인 $\hat{x}(t+f)$ 을 추정함에 있어 불확실성을 최소화하는 문제이다. 결국 예측 문제는 불확실성을 최소화하기 위해 상태진화의 전환점에서 방향성(directionality)과 변화량(variance)을 사전에 추정하는 문제라 할 수 있다.

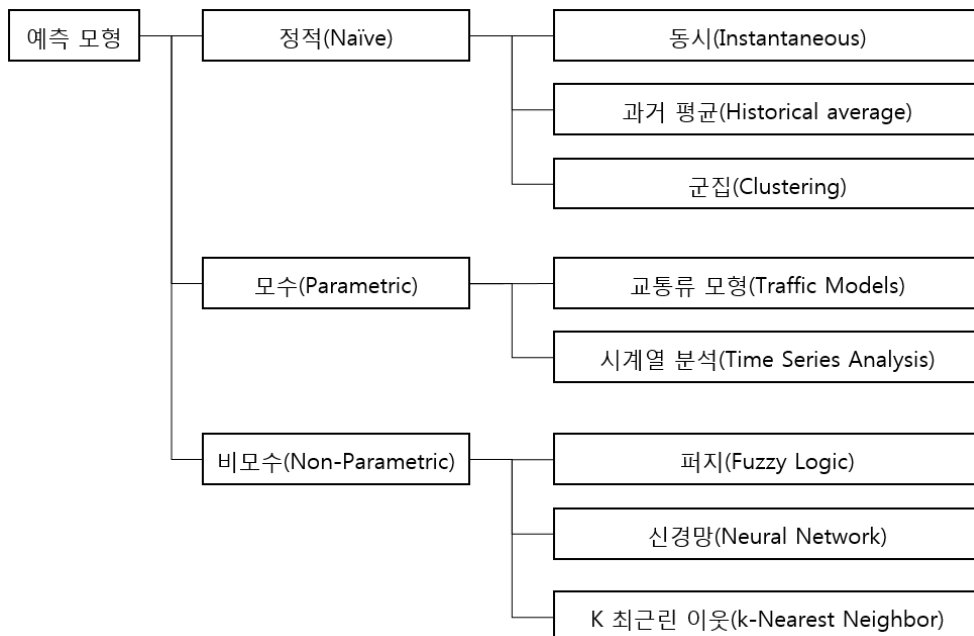
Given $X_c(t)$ with l 식(2-3)

Predict $\hat{x}(t+f)$ with min. of uncertainty, $\forall f, f \in F$

일반적으로 예측문제에서 변수의 차원은 상태벡터가 한 개인 경우 단일 변수(Univariate), 한 개 이상인 경우 다중 변수(Multi-variate)로 구분된다. 그리고 입·출력 변수의 관계는 [단일 변수 - 단일 변수], [다중 변수 - 단일 변수], [다중 변수 - 다중 변수]이며, 일반적으로 [단일 변수 - 단일 변수], [다중 변수 - 단일 변수]의 관계가 주를 이룬다.

4) 예측모형의 분류

예측문제를 해결하고자 하는 일련의 노력은 예측모형을 통해 수행되며, 다양한 예측모형이 제안 및 고도화되고 있다. <그림 2-4>는 예측 문제의 해결을 위해 주로 이용되는 예측모형의 분류를 보여주고 있으며, 정적기법(Naive)과 동적 모형인 모수(Parametric)와 비모수(Non-parametric) 모형으로 구분할 수 있다.



<그림 2-4> 예측모형의 분류

5) 정적(Naive) 기법

정적기법은 현장에서 널리 적용되는 방법이지만, 현행 상태의 변동을 고려하지 않은 방법으로서 모형에 대한 가정을 하지 않는다. 정적기법으로는 동시(Instantaneous) 기법, 과거평균(Historical average) 기법, 군집(Clustering) 기법 등이 주로 이용된다.

동시기법은 주로 동시 통행시간(Instantaneous travel time, 이하 ITT)의 산정을 위해 이용되며, 장래에도 현행상태가 지속된다고 가정한다. 연산이 매우 빠르며 현장 적용성이 뛰어나지만 장래 상태가 (급격히) 변화하는 경우 예측 성능은 매우 낮다(Lint et al., 2002; Park and Rilett, 1998; Park et al., 1999; Rilett and Park, 2001).

과거 속도 Profile을 이용하는 과거평균기법은 고도화된 모형에 비하여 예측력은 낮지만, 단기 예측기법의 예측영역을 초과하는 다중시간대 예측의 경우 반복적 행태가 나타나는 도로구간에 대하여 타 모형에 비하여 우수한 성능을 보일 수 있다(Kamarianakis and Prastacos, 2003; Park and Rilett, 1998; Park et al., 1999; Rilett and Park, 2001; Wu et al., 2004). 동시기법과 과거평균 기법의 조합은 보다 긴 예측 시간대를 예측할 수 있으나 높은 예측력을 보이지 않는다(Park and Rilett, 1999). 그리고 반복적 특성이 강한 도시부의 연속류 도로구간에 적합하며 비반복적 특성이 강한 지역간 연속류 구간의 경우 예측력은 낮은 편이다.

군집기법은 유사 교통 패턴을 기반으로 요일, 시간대에 대한 특정 그룹의 평균값을 적용하는 방법이며, 예측모형의 입력자료를 전처리하기 위해 이용되기도 한다(Chrobok et al. 2004). 군집기법의 경우 과거평균기법이나 선형회귀 모형보다 우수한 성능을 보이는 경우가 많다.

이상에서 고찰한 Naive 기법은 모수 및 비모적 모형에 비하여 낮은 예측력을 보인다. 그러나 현재로는 실제 적용성에 있어 가장 보편적인 실행가능한 방안중 하나이며, 특히 중·장거리 통행시간의 정보제공에 있어 유력한 대안 중 하나라 할 수 있다. 이는 중·장거리 통행시간의

예측에 적용할 수 있을 정도로 단기예측모형의 시간적 예측영역이 수용할 만한 수준에서 확장되지 않았기 때문이다.

6) 모수(Parametric) 모형

모수모형(Parametric model)은 교통류 모의실험 모형과 통계기법을 기반으로 하는 시계열 분석(Time series analysis) 기법으로 구분할 수 있다. 교통류 모의실험 모형은 교통류 행태에 대한 지식을 모형의 구조에 포함시켜 장래 상태를 추정한다. 교통류 모의실험 모형은 교통사고와 같은 비 반복적 행태를 평가할 수 있는 장점이 있어 주로 동적 교통관리의 의사결정을 위한 수단으로 이용된다. 일반적으로 교통류 모의실험 모형은 “Wardrop’s First Principle”에 의한 네트워크 균형의 개념을 이용한 기종점(Origin-destination, 이하 OD) 통행량을 통행배정한 교통수요를 이용한다. 따라서 교통류 모의실험 모형의 예측문제는 동적 기종점 통행량 예측으로 귀결된다.

교통류 모의실험 모형은 거시적(Macroscopic), 중시적(Mesososcopic), 미시적(Microscopic), 그리고 Kerner의 3현시 교통류 모형으로 구분할 수 있다. 거시적 모형은 거시적 교통류 변수인 밀도, 평균속도, 그리고 교통량이 이용된다(Cremer, 1995). 교통수요를 네트워크에 부하>Loading)하기 위한 통행배정기법은 정적 통행배정과 동적 통행배정이 이용된다. 정적 통행배정 모형은 전체 OD 통행량의 균형상태만 고려된다. 따라서 출발 시간을 고려하지 않으며, 앞막힘(Spillback) 현상 등으로 인한 경로우회 등을 설명할 수 없다. 이러한 정적 통행배정의 단점에도 불구하고, 모형이 심플하고 연산속도가 타 모의실험 모형에 비하여 빠르기 때문에 현장에서 많이 적용되고 있다. 동적 통행배정 모형은 시간에 따른 수요의 변동을 고려하며, 유체역학을 기초로한 거시적 교통류 해석 모형을 이용하여 시간에 따른 교통량과 밀도의 전과과정을 해석한다(Hounsell and Ishtiaq, 1997; Lin et al. 2004).

미시적 모형은 도로망에서 개별 차량간의 상호작용을 고려하면서 모든

개별 차량을 모의실험함으로써 교통류를 해석한다. 따라서 미시적 모의 실험 모형은 연산에 많은 시간이 소요된다. Nagel과 Schreckenberg(1992)에 의해 개발된 Cellular Automata(CA) 차량모형을 기초로한 다양한 차량모형이 모의실험에 적용되었다. CA기반 교통류 모형은 타 미시적 모형에 비하여 연산수행속도가 실시간 적용이 가능한 수준으로 매우 빠르며, 차량의 가다서다 현상과 동시 교통류를 설명할 수 있는 장점이 있다. Dailey와 Taiyab는 OD통행량을 이용하여 장래상태를 예측하였으며, Chrobok et al.(2004)는 회전비를 이용하여 장래상태를 예측하였다. 장현호(2002)는 CA차량모형을 이용하여 단속류의 장래상태를 예측하였으며, CA차량모형 기반의 장래 교통상태 예측은 교통사고와 같은 비반복적 교통상태에서 장래 교통상태의 분석에 적용되었다(Chang et al., 2007; Kim et al., 2008; Shah et al., 2008). 현재 미시적 모형을 이용한 대규모 고속도로망 시뮬레이션은 CA모형을 기반으로 하는 OLSIM(On-line Simulation) 정도가 실시간으로 적용되고 있는 수준이다(<http://www.autobahn.nrw.de>).

중시적 모형은 거시적 모형과 미시적 모형의 결합으로서 거시적 이론을 이용하여 통행배정을 수행한 후, 개별 차량은 산정된 거시적 교통류 변수에 기초하여 네트워크에 부하된다. 중시적 모형의 최대 장점은 대기행렬을 구현할 수 있다는 것이나, 미시적 모형 수준에서 개별차량의 행태를 설명할 수 없다. 최초의 중시적 모형으로 Van Aerde(1996)에 의해 개발된 Integration이 소개되었다. 대표적 중시적 모형으로는 DynaMIT이 있으며, 이력과 실시간 정보를 이용하여 OD 통행량을 추정하고 중시적 차량모형을 이용하여 도로망에 차량을 부하한다(Ben-Akiva et al. 2001).

이상의 모의실험 기반 모형들은 교통사고 등의 비 반복적 상태와 다양한 교통제어 전략을 포함하는 동적 교통류 관리에 적합하나, 교통류 및 차량 모형의 파라미터 정산이 용이하지 않으며 예측의 중심을 상태가 아닌 동적 OD추정에 둔다. 현재까지 동적 OD추정은 지속적인 연구분야이며, 모형의 연산수행 속도가 타 미 차량모형에 비하여 상대적으로 느리다는 단점이 있다.

시계열 분석(Time series analysis, 이하 TSA) 기법은 과거 관측값과 오차항(Error term)의 함수를 이용하여 예측 변수의 값을 추정하는 과정으로서 정적상태를 가정한다. 선형회귀(Linear regression, 이하 LR), 국부가중회귀(Locally weighted regression, 이하 LWR), ARIMA, 칼만필터링(Kalman filtering, 이하 KF), 가우시안 최우수 추정(Gaussian Maximum Likelihood, 이하 GML) 등 다양한 TSA 분석기법이 제시되었다.

LR모형은 공변량의 선형조합을 가정한 예측함수(Forecasting function)이다. 모형이 간단하기 때문에 연산수행속도가 빠르며, 만족할 만한 정확도를 보이는 것으로 보고되고 있다(Chrobok et al., 2004; Kwon et al., 2000; Lan and Miaou, 1999; Lingras et al., 2002; Rice and Zwet, 2004; Sun et al., 2003; Zhang and Rice, 2003; Zhong et al., 2005).

LWR모형은 국부회귀(Local regression)을 이용하며, 각 예측시점의 예측오차는 현행 관측값에 가깝게 가중된다. LWR모형은 LR모형에 비하여 우수한 결과를 보였으며(Sun et al., 2003; Zhong et al., 2005), 연산 수행시간도 매우 짧은 장점이 있다.

ARIMA 계열모형은 가장 일반적인 통계반 TSA 기법이며, Box-Jenkins(1976) 모형을 기반으로 한다. ARIMA모형의 예측력을 향상시키기 위하여 다양한 형태의 ARIMA 계열모형 [Seasonal ARIMA(Guo, 2005; Smith et al., 2003; Williams et al., 1998), subset ARIMA(Lee and Fambro, 1999), Kohonen ARIMA(Voort et al., 1996), ARIMAX(Williams, 1999), VARMA와 STARMA(Kamarianakis and Prastacos, 2003), 지수평활화(Crobok et al., 2004; Part et al., 1998)]이 제시되었다. ARIMA 계열모형은 서로간에 연구결과가 달라 어떤 계열의 모형이 예측 정확도가 우수한지에 대한 논란이 있으나, Seasonal ARIMA가 타 ARIMA 계열모형에 비하여 우수하다는 연구결과가 주를 이루고 있다.

KF모형은 Kalman(1960)에 의해 개발되었으며, 현행 상태벡터와 바로 직전 상태, 그리고 백색소음과 오차의 잔차를 이용하여 장래 상태를 예측한다. KF모형은 연산수행속도가 매우 빠르고 시계열의 변동을 효과적

으로 설명하기 때문에 다수의 폭넓은 연구분야와 실제 ITS 시스템에 탑재/운영되고 있다(Ben-Akiva et al., 1992; Chang et al., 2012b; Chen and Grant-Muller, 2001; Chien and Kuchipudi, 2003; Kuchipudi and Chien, 2003; Okutani and Stephanedes, 1984; Whittaker et al., 1997; Yang, 2005; Yoon and Chang, 2014).

GML모형은 “과거 평균에서 가능한 적게 예측값은 벗어하며, 예측의 증가는 과거의 증가에서 가능한 적게 벗어난다”는 2가지 가정을 기반으로 개발되었다. 예측력은 타 TSA기법보다 우수한 것으로 보고되었다(Nicholson and Swann, 1974).

이상의 확률적 상태이론을 기반으로 하는 TSA기법들은 시계열의 모델링에 있어 정적(Stationary) 상태를 가정하지만 실제 교통류 시스템은 매우 동적인 시스템이다. 일반적으로 TSA기법은 계절성(Seasonality)이 필요하며, 교통류 이론보다는 통계적 함수에 의존한다. 특히, ARIMA 계열모형은 평균값에 의존하는 모형의 본질적 특성으로 인하여 상태변화가 안정정일 때 예측력은 우수한 반면, 상태의 전환점 이후 상태를 예측하지 못 하는 한계가 있다(Vlahogianni et al., 2005). 따라서 상태가 급격하게 변하는 단속류의 행태를 설명하지 못 한다(Chang et al., 2012b; Yoon and Chang, 2014).

7) 비모수(Non-parametric) 모형

비모수(Non-parametric) 모형은 변수(Parameter)가 결여됨을 의미하지 않으며, 파라미터의 개수와 내재적 특성(Nature)이 가변적이며 고정되지 않음을 의미한다. 비모수 모형의 구조 및 파라미터는 데이터 즉, 예측대상 시스템의 상태변화에 따라 결정되며, 수학적 방정식이 아닌 경험적 지식 기반의 문제해결 접근법이다. 따라서 TSA기법과 같은 모수적 접근법에 비하여 많은 양의 데이터가 요구되지만, 시스템 상태의 시간적 변화에 대한 깊은 이해와 지식을 필요로 하지 않으며 알려지지 않은 파라미터(Unknown parameter)와 특성으로 설명되는 내재된 행태를 동적으로 설

명할 수 있다는 장점이 있다. 비모수 모형은 퍼지(Fuzzy), 신경망(Neural Network, 이하 NN) 그리고 k -최근린 이웃(k -nearest neighbors, 이하 KNN) 등으로 구분할 수 있다.

퍼지모형은 IF-THEN rule의 집합으로 구성되는 의사결정 논리라고 할 수 있다. 현행 상태는 하나 또는 여러 개의 If에 해당하며, 장래 상태는 THEN에 의하여 예측된다. 퍼지로직 기반의 예측모형은 만족할 수준의 결과를 도출하였지만(Coufal and Turumen, 2004; Li et al., 2006), NN모형과 결합됨으로써 더 우수한 예측력을 보이는 것으로 보고되었다(Huisken, 2003). 예측분야에서 퍼지로직은 NN과 같은 모형과 결합되어 이용되고 있으며, 퍼지로직 자체 만으로 예측 문제를 해결하는데 한계가 있다.

NN모형은 비선형적인 교통상태의 변화과정(Non-linear process)을 모델링할 수 있기 때문에 ARIMA, KF 기법과 더불어 교통상태 예측문제에 가장 널리 연구된 모형이라 할 수 있다. NN모형은 NN의 최적화과정 즉 학습과정(Training process)을 통해 현행 상태를 신경망에 학습시킨 후 학습된 지식을 이용하여 장래 상태를 추정하게 된다. NN모형은 은닉층의 연결강도로 구성되는 인공지능의 내부를 이용자가 볼 수 없기 때문에 이른바 “Black box” 알고리즘으로 불리기도 한다. NN의 구조, (효과적인 훈련과정을 통한) 학습, 그리고 (NN 알고리즘의 훈련과정에서 요구되는) 연산시간의 단축 등을 통하여 예측 정확도와 적용성을 향상시키기 위한 일련의 노력으로 다양한 형태의 NN모형들이 제시되었다. NN기반 모형은 전통적인 NN(Traditional Neural Network, 이하 TNN), 학습과정, 신경망의 내부 구조, 입력자료의 전처리, 시공간(Temporal and spatial) 패턴으로 구분할 수 있다.

TNN은 1개 입력층(Input layer), 복수개의 은닉층(Hidden layer), 그리고 1개의 출력층(Output layer)으로 구성되며, 입력층은 자극, 은닉층은 판단, 그리고 출력층은 반응의 역할을 수행한다. 각각의 층은 복수개의 뉴런(Neuron, 신경계를 이루는 기본적인 단위 세포)으로 구성되며, 층간은 해당 층에 속하는 뉴런간의 연결강도로 구성된다. 전형적인 NN의 학습과정은 연결강도(0~1.0)를 조정하는 과정으로서 출력층의 오차를 신경

망의 후방 즉, 판단과 자극의 방향으로 책임할당을 통해 전파시키는 오류 역전파(Error backpropagation) 학습법이 주로 이용되며, 오차범위에 수렴할때가 반복된다. 따라서 TNN은 일반적으로 BPNN(Back propagation neural network)로 불리우며, 교통변수 예측에 있어 많은 연구가 보고되었다(Dougherty and Cobbett, 1997; Ishak and Alecsandru, 2004; Kisgyorgy and Rilett, 2002; Ledoux, 1997; Park and Rilett, 1999; Zhang, 2000).

NN모형의 장점은 학습과정을 통해 상태의 진화에 내재된 지식을 NN이 습득하고 반응한다는 것이다. 반면 단점은 학습과정에서 불가피하게 수반되는 증가된 연산이며, 이는 실제 적용에 있어 장애요인이다. 이러한 장단점에 대해 장점을 강화하고 단점을 최소화하기 위한 다양한 학습과정 NN이 제시되었다. 널리 이용되는 학습과정은 결합 경사 알고리즘(Conjugate gradient algorithm)이며, BPNN과 유사한 결과를 보였다(Innamaa, 2005). 진화적 학습을 이용하는 진화 신경망(Evolutionary NN, 이하 ENN)은 BPNN보다 빠른 학습시간과 높은 예측 정확도를 보였으며(Abdulhai et al., 1999; Lingras et al., 2002; Zhong et al., 2005), 진화의 과정은 주로 유전자(Genetic) 알고리즘이 이용되었다.

NN의 효과적인 구조를 통한 노력은 분할정복(Divide-and-conquer) 전략(Ishak and Alecsandru, 2004)에 기반한 학습단위의 NN(Modular NN, 이하 MNN), 방사 기저 주파수(Radial Basis Frequency NN, 이하 RBFNN), Fuzzy-NN 등을 들 수 있다. MNN의 입력값은 다수의 하위 망(Subnetwork)로 구성되며, 학습속도 및 예측력은 높게 나타났다(Ledoux, 1997; Park and Rilett, 1998; Zhang et al., 2000; Zheng et al., 2006). RBFNN은 은닉층의 중심 뉴런과 입력 상태벡터간의 상태간 거리를 이용하며(Amin et al., 1998; Huisken, 2003; Park et al., 1998; Xie and Zhang, 2006), 예측 정확도 측면에서 BPNN과 유사한 결과를 보였다. Fuzzy-NN은 은닉층에 퍼지규칙을 추가한 NN이며, 상태의 변동이 심한 경우 높은 예측 정확도를 보였다(Coufal and Turumen, 2004; Stathopoulos et al., 2008; Yin et al., 2002).

일반적으로 시계열 교통자료는 소음(Noise)이 포함되어 있다. 소음이

포함된 입력자료는 NN의 은닉층에서 잘못된 자극으로 인식되며, 이는 예측 실패로 이어질 수 있다. 따라서 입력자료의 소음을 제거한 상태를 입력자료로 적용하기 위한 입력자료 전처리와 NN이 결합된 모형에 관한 연구가 수행되어 왔으며, 대표적인 모형으로 스펙트럴 NN(Spectral NN, 이하 SNN)과 파형요소 NN(Wavelet NN이하 WNN)이 있다. SNN은 푸리에 변환(Fourier transformation)을 이용한 선형으로 구분된 상태로 변환한 후 이를 BPNN의 입력값으로 이용한다. 특히, SNN모형은 다중 시간대의 예측에 있어 타 NN모형에 비해 우수한 정확도를 보였다(Park et al., 1999; Rilett and Park, 2001). WNN은 파형요소 변환(Wavelet transformation)을 이용하여 입력자료의 소음을 제거하고, BPNN의 Sigmoid 함수대신 이용하게 된다(Xie and Zhang, 2006). WNN은 연산수행 속도와 예측 정확도 측면에서 BPNN보다 우수한 것으로 보고되었다(Jiang and Adeli, 2005; Xie and Zhang, 2006).

시공간 패턴은 순환 NN(Recurrent NN, 이하 RNN)에서 이용된다. RNN은 별도의 메모리 공간에 은닉층의 출력값을 저장한다. 따라서 저장된 이전 예측 시간대의 출력값은 순환 패턴의 인지 메커니즘으로 이용된다(Ishak et al., 2004). 상태공간(State space)을 이용하는 SSNN은 RNN을 기본 골격으로 하며(Lint et al., 2002), SSNN에서 은닉층의 뉴런은 전체 경로를 구성하는 도로구간으로 해석되어 진다. 따라서 은닉 뉴런의 가중치는 구간통행시간으로 해석된다. SSNN은 Naive기법보다 매우 우수한 예측 정확도를 보였지만, 다른 NN모형과는 비교되지 않았다(Lint and Schreuder, 2006)

이상에서 고찰한 NN 계열모형은 폭 넓고 깊은 연구에도 불구하고, 한편의 논문(Park and Rilett, 1998)을 제외하면 실제 도로망에 적용되지 않았다. 이는 대다수의 연구가 단일 지점 또는 단일 구간의 자료를 이용하여 수행되었기 때문이며, ITS 시스템 시간의 견지에서 볼 때 NN의 연산 수행속도는 동적 교통류 관리 및 정보제공에 불리하기 때문이다.

TSA기법과 NN모형과 달리 KNN-NPR 모형은 현행상태와 유사한 과거상태를 탐색하는 과정을 통해 현행상태와 유사한 k 개의 과거상태인

KNN을 구축한 후, 직접평균 또는 상태간 거리의 역수로 가중한 평균과 같은 예측식을 이용하여 장래상태를 예측한다(Chang et al., 2012b).

KNN-NPR 기반의 예측 연구들은 Naive 예측기법보다는 우수한 성능을 보였으며, SARIMA와 BPNN과 유사한 성능을 보였다(Smith and Demetsky, 1996; Smith et al., 2003; Rice and Zwet, 2004). KNN-NPR 모형은 교통류의 상태 진화에 대한 이해를 방대한 양의 과거 상태에 전적으로 의존한다. 즉, 과거 자료의 다양성과 질에 전적으로 의존한다(Chang et al., 2012_{a,b}). 그러나 몇몇의 연구는 KNN-NPR모형의 데이터 요구수준을 충족하지 않은 상태에서 KNN-NPR모형을 예측성능이 낮은 Benchmark모형으로 선정하고 연구의 결과를 도출한 경우가 많다는 지적이 있다(Yoon and Chang, 2014). 이와 반대로 Smith and Demetsky (1997)의 연구에서는 NN, ARIMA보다 KNN-NPR의 예측력이 우수하게 나타났다.

최근들어 ITS의 자료수집 기술이 발달하고, 대용량 자료를 지원하게 되면서 KNN-NPR을 이용한 예측연구가 증가하고 있다(Chang et al., 2012a). Chang et al.(2010)과 Kim et al.(2011)은 경로통행시간의 예측에 우수한 예측 정확도를 보였으며, Chang et al.(2012_{a,b})는 교통량의 예측과 추정에 있어 다중시간대 예측에도 불구하고 SARIMA, KF기법의 단일 시간대 예측의 정확도 보다 매우 우수한 성능을 보였다.

예측영역의 확장측면에서 KNN-NPR은 타 접근법에 비하여 장점을 가진다. 기존의 단기예측의 모형들은 수용할만한 예측오차의 수준에서 예측영역 확장에 실패한 반면, KNN-NPR 기반 예측연구의 경우 예측영역의 확장에 성공한 사례가 보고되고 있다(Smith and Demetsky, 1996; Chang et al., 2010, 2012b). 추가로 시계열자료의 변동이 극심한 단속류의 상태 예측에 있어 SARIMA와 KF보다 매우 우수한 성능을 보이는 것으로 보고되었다(Yoon and Chang, 2014). 사실상 극심한 변동상태의 예측에 있어 ARIMA와 KF 기법은 방향성과 변화량에 대하여 완전 예측실패를 보이는 경우가 많다.

8) 기존 교통상태 예측 연구 고찰의 소결

교통류 시스템의 완벽한 상태이론이 정립되지 않은 상태에서 모형 개발자간의 상반되는 주장과 교통류가 동적 시스템이라는 공통된 주장속에 다양한 예측모형이 제시되었다. 기존연구 고찰 결과를 도로망 소통상태와 중·장거리 통행시간 예측의 관점에서 정리하면 다음과 같다.

첫째, ITS분야에서 예측정보의 제공과 활용은 일반적으로 5~15분 이내의 단기 목적으로 수행되어 왔기 때문에 대다수 ITS 예측연구는 예측영역을 단기로 한정하고 있다. 따라서 예측영역을 초과한 장래상태가 예측영역의 마지막 상태와 동일하게 15분 동안 지속된다고 가정하더라도 30분 이상의 중·장거리 통행시간에 비하여 시간적 영역이 짧다.

둘째, Naive기법을 포함한 대부분의 TSA 계열기법과 NN 계열모형은 단일 지점 또는 구간에 대하여 [단일변수 또는 다중변수]→[단일변수]입·출력이라는 구조적 문제가 있기 때문에 중·장거리 통행시간의 예측에 적합하지 않다. 따라서 중·장거리 통행시간에 적합하도록 예측영역이 확장된 장래 소통상태를 예측하기 위해서는 [도로망의 모든 구간에 대한 단일 또는 다중변수]→[도로망의 모든 구간에 대한 시간대별 통행시간]과 같은 입·출력 구조의 변경이 필요하다.

셋째, 모의실험기반 예측모형을 제외하면 대부분의 예측모형은 복수개의 구간으로 구성되는 도로망에서 교통류 상태진화의 시·공간적 메커니즘이 결여되어 있다. 따라서 예측영역이 확장될 경우 도로망의 차원에서 혼잡의 발생→진화→성숙→완화→소멸의 과정을 설명할 수 없다. 그리고 모의실험 예측모형의 경우, 소수의 온라인 시뮬레이터 관련연구(Chang et al., 2007; Chrobok et al., 2004; Shah et al., 2008)를 제외하면 대규모 도로망의 실시간 모의실험에 대한 연구사례가 보고되고 있지 않으며, 동적 OD를 대규모 가로망으로 확대하지 못 하고 있는 실정이다.

넷째, 기존의 대다수 연구들은 자료수집의 한계로 인하여 소량의 정상자료를 이용하여 연구를 수행하였다. 실제 ITS 데이터베이스는 통신장애, 자료수집단의 고장 및 일시적 장애, 시스템 유지보수 등 다양한 원인으로 결측자료를 포

함하고 있다. 그러나 현행 자료뿐만 아니라 과거 자료를 입력변수로 이용하는 모든 예측모형에서 누락자료에 대한 부분을 고려하고 있지 않다. 추가로 적용 도로망의 도로구간 개수와 입력자료의 기간이 증가할수록 누락자료의 양은 급격히 증가한다. 따라서 예측모형을 대규모 도로망에 적용하기 위해서는 누락자료를 예측모형이 수행과정에서 자동으로 처리해야 한다.

다섯째, 오늘날의 교통자료 수집체계는 기존의 지점검지체계에서 개별 차량을 수집단으로 하는 구간검지체계로 변화하고 있다. 구간검지체계는 수집자료의 정확성과 수집단의 유지·보수 측면에서 장점이 있지만 교통자료의 수집에 시간처짐(Time lag)이 발생한다. 예를 들어 차량이 3km 도로구간을 10km/시로 주행하였다면, 자료의 수집에 소요되는 시간은 0.3시간(1080초)이다. 자료의 집계시간 길이와 예측을 위한 시간대 길이를 5분으로 가정하면, 5개 시간대($4.6=1080\text{초}/300\text{초}+1$)에 해당하는 시간처짐이 발생한다. 따라서 시스템의 기준에서 시간대($t+1$)의 실시간 정보를 제공하기 위해서는 6개 시간대 이후 즉, 30분 이후를 예측해야 한다. 특히, 이러한 혼잡상황의 경우 정보의 가치는 비혼잡시 보다 높다. 그러나 기존에 보고된 예측모형은 장래 30분 이후까지 예측영역을 확장하지 못하고 있다고 판단된다. 추가적으로 이를 극복하기 위하여 시간대의 길이를 연장하게 되면 장래 상태의 불확실성 또한 증가함으로 되어 결국 예측실패를 피할 수는 없다(Chang et al., 2010).

여섯째, ITS 시스템에 예측모형을 탑재하기 위해서는 정보제공 및 교통류 제어전략과 시스템이 요구하는 모형의 연산수행속도를 충족해야 한다. 허용 예측오차의 범위내에서 현재의 단기예측모형이 예측영역이 확장되었다고 가정하고, 도로구간의 개수가 100개인 도로망의 현재 시각부터 향후 6시간까지 5분 시간간격으로 구간 통행시간을 예측한다고 가정하면, 총 7,200 [= 100개 도로구간 × 6시간 × 12개 시간대]번의 예측과정을 반복해야 한다. ITS 시스템의 입장에서 탑재된 예측모형의 연산 수행시간(ϵ)은 시간대가 t 에서 $t+1$ 로 변화하는 예측시점(t_f)에서 $t_f+\epsilon$ 까지이며, $t_f/(t_f+\epsilon) \rightarrow 1.0$ 에 가까워야 한다. ϵ 이 증가할수록 동적정보의 생애주기(Life cycle)는 짧아지며 실시간 및 사전대응의 기회는 적어지기 때

문이다. 이는 고도화된 예측모형인 SARIMA나 NN모형이 실제 시스템에 탑재되지 못하는 근본적인 원인이기도 하다. Smith et al.(1999)가 지적한 바와같이 SARIMA의 연산수행속도는 시스템 시간의 측면에서 6박 7일과 같다고 지적하였으며, NN계열의 모형이 실제 대규모 도로망에 적용되지 못한 원인 중의 하나이기도 하다. 반면, KNN-NPR모형의 경우 Hadoop과 같은 대용량 처리 및 검색기술이 개발되면서 매우 빠른 연산속도로 시스템 요구시간을 충족하는 것으로 보고되고 있다(Chang et al., 2010; Smith and Oswald, 2003).

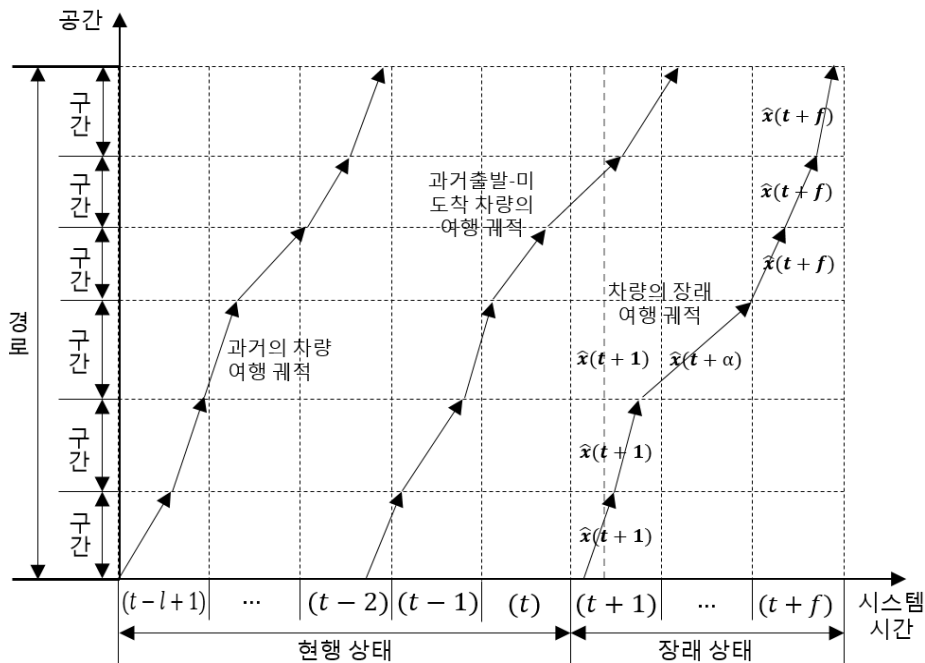
일곱째, 예측 정확도를 향상시키기 위한 모형의 고도화는 실제 ITS 시스템 탑재시 또다른 장애요인이 되고 있으며, 첨단화된 예측모형이 극복해야 할 또 다른 과제가 되고있다(Chang et al., 2012_b). 일반적으로 ITS 시스템 유지관리 요원 및 실무 담당자의 경우, 시스템에서 유지·관리하는 데이터의 구조에는 폭넓은 현장지식을 소유하고 있지만, 고도화된 모형의 구조, 알고리즘 그리고 파라미터의 최적화 등에 대한 전문 지식을 보유하지 않은 경우가 많다. 이러한 원인으로 시스템에 탑재되어 운영중인 모형의 구조변경과 파라미터 재정산시 상당한 시간이 요구되는 문제가 발생하며, 이는 사업의 측면에서 시간과 예산의 문제로 귀결되는 경우가 많다. 따라서 수용 가능한 오차수준에서 모형의 정확도가 확보되었다면, 예측모형의 구조변경 및 파라미터 재정산 등에 대한 운영적 측면을 고려해야 한다.

마지막으로 van Hinsbergen et al.(2007)이 지적한 바와 같이 특정 모형의 장래상태 설명력이 뛰어나게 우수하지 않으며, “using the right prediction method(s) at the right time”라고 지적하였다. 또한 21세기 ITS분야 예측연구는 연속류의 경우 1시간 이후까지의 다중시간대 예측과 단속류의 경우 단일 시간대에서 다중시간대에 걸친 수용 가능한 오차수준에서의 예측을 통해 기존의 실시간 정보제공 및 제어에서 보다 적극적인 사전대응(Proactive) 정보제공 및 제어를 지원함으로써 ITS의 편익을 극대화해야 한다(Chang et al., 2012_b; Yoon and Chang, 2014)고 지적하였다.

3. 구간기반 통행시간 예측모형

1) 구간기반 통행시간 예측의 시·공간적 구조

일반적으로 시간대(t)에서 출발하는 통행의 출발지(o)에서 도착지(d)까지 경로통행시간, $PTT_{o,d}(t)$ 은 일련의 도로구간을 시간에 따라 여행하는 시·공간상에서 차량의 이동궤적으로 설명되며, <그림 2-5>와 같다. 구간기반 경로통행시간(Link-based path travel time, 이하 LPTT) 예측모형은 식(2-3)의 예측문제 풀이를 통해 구간별 장래 대표 통행시간을 예측하고, 예측된 일련의 구간통행시간을 이용하여, 대표 경로통행시간을 산정한다. 따라서 t 에서 $o \rightarrow d$ 의 $PTT_{o,d}(t)$ 를 예측하기 위해서는 시스템 시간 기준으로 모든 장래 구간통행시간 $\hat{x}(t+f)$ 를 예측해야 한다. 다시말해 물리학적으로 $PT_{o,d}(t) \leq TL \cdot F$ 의 조건을 만족해야 한다. 여기서 $f \in F$, $F = [1, 2, \dots, f]$ 는 시간대 길이(TL)로 분할되는 예측의 영역이다.



<그림 2-5> 구간기반 통행시간의 시·공간 예측 영역

2) 구간통행시간 예측 연구

<그림 2-3>에서 시스템 시간을 기준으로 실제 시각(t_{nw})과 가장 가까운 교통자료는 지점검지 체계의 경우 시스템과 수집단 간의 Call time을 고려하지 않을 경우 t 까지 수집되며, 구간검지 체계의 경우 자료의 수집에 발생하는 시간차집(τ)을 고려하면 $t-\tau$ 까지 수집된다. 그러므로 수집된 t 또는 $t-\tau$ 시간대의 자료는 t_{nw} 보다 과거의 교통류 상태이며, t_{nw} 기준의 실시간 통행시간 즉, 시스템 기준의 장래상태 통행시간 정보를 추정하기 위하여 구간통행시간 예측연구가 시작되었다.

지점검지체계를 기반으로 정보를 수집하던 초기 ITS에서 구간통행시간 예측연구의 예측영역(f)은 1로 설정되었으며, 장래 소통상태 정보라기 보다는 실시간 구간소통정보를 제공하기 위한 목적이다. 이후 차량 네비게이션의 등장에 따라 $o-d$ 간 경로통행시간 정보를 제공하게 됨으로써 구간기반 경로통행시간을 추정하기 위하여 $f > 1$ 이상으로 확장된 다중시간대 구간통행시간 예측연구가 진행되었다(Park and Rilett, 1998; Part et al., 1998).

구간통행시간의 예측에는 평활화, RL, LWR, ARIMA계열, KF 등의 전통적인 시계열 분석기법에서 NN까지 다양한 예측기법이 연구되었다. 그리고 2000년 이후 보다 정교한 모형들이 적용되고 있다. <표 2-1>은 2004~2013년까지 10년 동안의 구간통행시간 예측모형에 적용된 방법론의 시간대 길이, 예측 시간대의 개수, 독립변수, 수집자료를 요약해서 보여주고 있다. 주로 인공지능망, 통계기반의 기법, 베이지안, 그리고 Hybrid 모형이 주를 이루고 있다.

구간검지체계가 보편적으로 도입되면서, AVL 자료를 이용한 예측연구가 보고되었으며, 최근 이력자료의 양과 질이 향상되면서 KNN기반의 연구도 보고되었다. 시간대의 길이는 1~15분, 장예측 시간대의 개수는 1~4로 분석되었어 최대 예측영역은 15~20분으로 분석되었다. 그리고 예측영역은 2000년 이전의 단기예측의 수준과 동일하였다.

<표 2-1> 구간통행시간 예측연구의 고찰 결과

연 번	저자	도로 유형	시간대 길이(분)	예측영역 (개)	자료 수집	적용 모형
1	Lin et al.('04)	A	1	1	S	베이지안
2	Wu et al.('04)	M	3	1	D	통계기반
3	Innamaa('05)	M	1	1	D	신경망
4	Kwon & Petty('05)	M	15	1	D	통계기반
5	Oh et al.('05)	M	1	1	D	신경망
6	van Lint & van Zulen('05)	M	1	1	D	신경망
7	van Lint et al. ('05)	M	1	1	D	신경망
8	Dion & Rakha('06)	M	1	2	A	통계기반
9	Innamaa('06)	M	1	1	D	신경망
10	van Lint('06)	M	1	1	D	신경망
11	Wang et al.('06)	M	1	20	S	통계기반
12	Juri et al.('07)	M	1	1	S	통계기반
13	Li('08)	M	5	1	D/A	베이지안
14	van Lint('08)	M	1	1	D	Hybrid
15	Innamaa('09)	M	5	4	D	신경망
16	Jintanakul et al.('09)	M	5	1	S	베이지안
17	van Hinsbergen et al.('09)	M	5	3	D	신경망
18	Zou et al.('09)	M	5	1	D	신경망
19	Innamaa('09)	M	1	1	D	신경망
20	Yang et al.('10)	M	15	1	A	통계기반
21	Abu-Lebdeh & Singh ('11)	A	5	1	S	Hybrid
22	Bustillos & Chiu('11)	M	15	1	S	통계기반
23	Fei et al.('11)	M	1	1	D	베이지안
24	Khosravi et al.('11)	M	5	1	D	Hybrid
25	Oh and Park('11)	M	1	1	A	신경망
26	Simroth & Zähle('11)	M	1	1	GPS	통계기반
27	Soriguera & Robusté('11)	M	5	1	D/A	통계기반
28	Xia et al.('11)	M	15	1	D	통계기반
39	Wang and Shi('12)	M	1	1	D	베이지안
30	Mu et al.('12)	M	1	1	S	통계기반
31	Du et al.('12)	M/A	2	1	S	통계기반
32	Haworth & Cheng('12)	M/A	5	1	A	통계기반
33	Khan('12)	M	5	1	G	베이지안
34	Lu('12)	M	5	1	D	베이지안
35	Ma et al.('12)	M	1	1	S	통계기반
36	Qiao et al.('12)	M	5	1	B	통계기반
37	Zheng & van Zuylen('12)	A	1	1	G	신경망

주1) M(고속도로), A(간선도로)

주2) D(검지기), A(AVL), G(GPS), S(모의실험)

현행의 구간통행시간 연구가 단일 시간대를 목표로 하는 배경은 다음과 같다. 초기 미국의 고속도로 교통관리시스템(Freeway Traffic Management System, 이하 FTMS)은 0.5 Mile($\approx 0.805\text{km}$)간격으로 지점검지기를 설치하였다. 혼잡시 속도가 10km/시의 일 경우 구간통행시간은 4.83($=60\text{분} \times 0.805\text{km}/10\text{kph}$)분이 소요됨으로 $TL=5\text{분}$ 일 경우 $f=1$ 이 설계되며, 예측된 $t+1$ 의 구간별 통행시간은 정보제공을 위하여 일련의 구간로 구성되는 Con-zone단위로 집계된다. 따라서 t_{nw} 의 기준에서 동시 통행시간의 개념으로 실시간 구간소통정보를 제공하기 위한 목적이라고 할 수 있다.

구간검지체계 하에서 AVL자료를 이용하는 통행시간 연구의 경우, 구간의 길이를 짧게하여 구간검지의 과정에서 발생하는 시간처짐(τ)을 고려하지 않거나 TL 을 연장하여 수행되었다. 그리고 TL 을 1분으로 짧게 한 경우는 빠른 교통제어의 목적과 더불어 보다 빠른 동적 정보제공의 목표를 달성하기 위함이다. 추가적으로 TL 을 짧게 하여 교통정보를 제공할 경우 정보이용자의 반응과 혼란에 대한 논란이 있는 상황이다.

구간통행시간 예측모형의 고도화에 따른 단일 시간대 또는 단기 통행시간 예측의 정확도는 2000년 이전에 비하여 향상된 것으로 보인다. 그러나 예측 정확도의 향상이 교통정보 수집기술의 발달에 따른 오차의 감소인지 모형 고도화에 따른 예측오차의 감소인지는 정확하지 않으며, 이는 van Hinsbergen et al.(2007)의 “단기 예측모형의 정확도는 크게 차이가 나지 않는다”라는 지적과 의미를 같이한다고 판단된다.

3) 구간통행시간 예측모형의 문제점

기존 연구의 고찰 결과, 2000년대의 구간통행시간 예측모형은 기존의 통행시간 모형에 비하여 고도화되었으며, 첨단화된 정보수집 단에서 수집된 보다 신뢰도 높은 통행시간 정보를 이용하였기 때문에 우수한 성능을 보인 것으로 판단된다. 그럼에도 불구하고 구간통행시간의 예측영역은 최대 15분 이하로 나타났다. 따라서 통행시간이 15분 보다 긴 중·장거리 통행시간 정보의 제공시 15분 이후의 통행시간 변화는 과거 통행시

간 자료의 평균값 또는 장래 15분의 통행시간이 15분 이후에도 지속적으로 유지된다는 가정하에 정보를 제공해야 한다. 이러한 원인에는 다음과 같은 복합적인 요인이 있다.

첫째, 해당 구간의 현행 또는 이력 자료를 이용하여 장래 상태의 불확실성을 극복하기에는 한계가 있다. 현행 상태벡터를 이용하여 추정된 상태의 진화가 사실상 상태의 변화시점 즉, 통행시간 상태의 전환 시점(Turning point)과 변화량을 정확하게 추정할 수 없기 때문이다. 엄밀히 말해, 현재까지 교통상태 예측분야에서 소수의 단기예측 연구(Chang et al., 2012b; Stathopoulos et al., 2008; Vlahogianni et al., 2005; Yoon and Chang, 2014)를 제외하면 예측모형을 이용한 동적 교통상태의 경계(Boundary condition)에 대한 사전(Proactive) 이해에는 실패하고 있는 실정이다. 추가로 교통상태 예측의 경우 일정 상태가 안정적으로 지속되는 상태 보다는 상황이 급변하는 즉, 정상상태→혼잡상태, 혼잡상태→정상상태로 변화하는 시점에서 예측이 중요하다.

둘째, 경로 통행시간의 경우 일련의 연속된 링크 통행시간으로 구성됨으로 개별 구간의 통행시간뿐만 아니라 도로망 전체의 소통상태 변화를 고려해야 한다. 그러나 구간 통행시간/통행속도 모형을 이용하여 도로망 혼잡의 발생부터 소멸까지 장래 소통상태를 설명할 수 없는 문제점이 있다. 구간통행시간 (또는 통행속도) 이외에 타 교통변수를 고려한다 해도 도로망에서 교통수요의 시·공간에 따른 변화량과 방향성을 고려하지 않았기 때문에 단기예측의 예측범위보다 긴 장래 상태의 불확실성을 극복할 수 없는 문제점이 있다.

마지막으로 고도화된 모형들은 정교한 방법론과 추가적인 파라미터를 이용하고 모형 개발자의 정확한 정산과정을 통해 모형의 성능을 극대화하는 경우가 많다. 결국 변수값의 Fitting과 Calibration에 의한 모형의 정교화는 실제 모형이 시스템에 탑재될 경우 그 성능을 발휘하기 어렵다는 다음과 같은 적용 한계를 가지고 있다(Smith and Oswald, 2003). 고도화된 예측모형은 Smith and Oswald(2003)가 지적한 바와 같이 실제

시스템에 탑재된 후 모형의 운영 및 관리에 있어 극복해야할 새로운 문제를 낳고 있다. 일반적으로 고도로 정교한 모형을 이해하고 이를 시스템에서 운영하기 위해서는 교통류에 대한 폭넓은 지식과 이해 그리고 예측문제 모델링에 대한 폭넓은 경험이 필요하다(Chang et al., 2012b). 그러나 ITS 시스템을 유지·관리하는 현장의 요원들은 예측모형의 모델링, 모형의 구조변경 그리고 입력 파라미터의 정산 등에 대한 충분한 지식을 보유하지 않은 경우가 많다. 따라서 시스템 유지·관리 요원들은 시스템에 탑재된 모형의 사소한 연산 문제, 모형의 입·출력 구조변경, 모형의 내부구조 변경, 현장조건 변경으로 인한 파라미터의 재정산 등의 문제에 봉착하게 된다(Chang et al., 2012b; Smith and Oswald, 2003).

4) 구간기반 경로통행시간 예측모형

구간기반 경로통행시간 예측모형(Link-based path travel time, 이하 LPTT)은 <그림 2-5>와 같이 예측시점(t_f)에서 예측영역(F)까지 예측된 구간별로 통행시간 $\hat{x}(t+f)$ 를 이용하여 시간대(t)에서 출발하는 차량의 $o \rightarrow d$ 간에 존재하는 일련의 구간통행시간을 누적하여 경로통행시간(Path travel time, $PTT_{o,d}(t)$)을 산출한다. 이러한 링크기반 통행시간 예측모형의 목표를 물리학적으로 충족하기 위해서는 식(2-4)의 조건을 충족해야 한다. 식(2-4)에서 TL 은 시간대의 길이이며, F 는 TL 의 개수로 설명되는 예측의 시간적 영역이다.

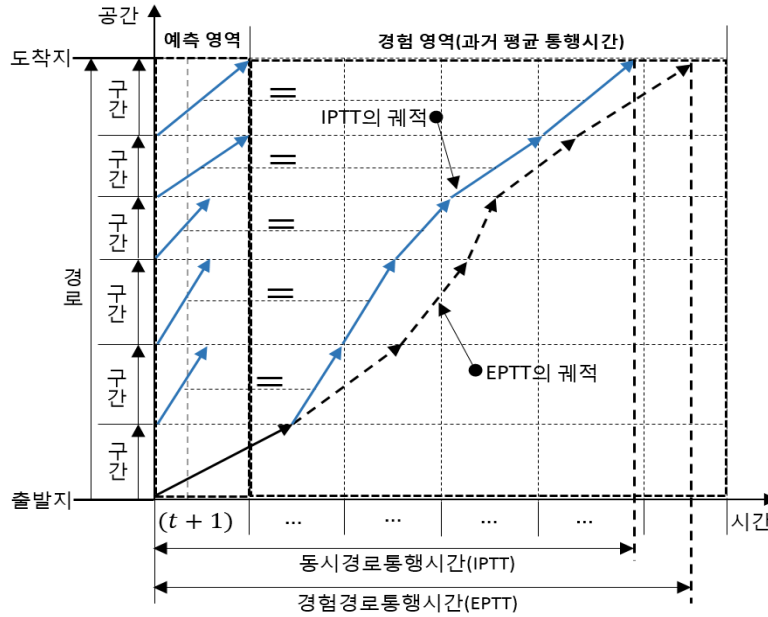
$$PTT_{o,d}(t) \leq TL \cdot F \quad \text{식(2-4)}$$

현재의 교통상태 예측모형으로 통행시간의 길이가 긴 중·장거리 통행의 구간기반 통행시간 예측에 있어 식(2-4)를 충족하기는 어려운 실정이라고 판단된다. 구간통행시간 연구에 대한 기존 연구의 고찰 결과, 보편적으로 f 는 1이며, TL 은 5분 이하로 설정하고 있다. TL 을 15분으로 연장하여 전체 예측영역의 길이를 확장하려는 노력도 있으나, TL 을 연장

할 경우 동적특성은 감소하게 된다. Park and Rilett(1998)의 연구에서 $f \geq 3$ 인 경우 예측오차는 10%이상으로 급격히 증가하는 것으로 보고되었다. 이후 통행시간 또는 속도를 예측하기 위한 다양한 기법을 이용한 연구가 보고되었으나 f 의 증가에 따른 예측오차의 급격한 증가, 즉 장래 불확실성을 극복하지는 못하고 있는 실정으로 판단된다. 일반적으로 TL 을 연장하거나 f 를 증가시킴과 동시에 장래 상태의 불확실성도 증가하게 됨으로 예측오차는 증가하게 된다(Chang et al. 2010). 이상에서 기술한 단기예측기법의 한계내에서 구간기반 경로통행시간 정보를 생성·제공하기 위한 대안으로 동시 또는 현행 경로 통행시간(Instantaneous or current path travel time, 이하 IPTT)과 경험 경로통행시간(Experienced-based path travel time, 이하 EPTT)이 주로 이용되고 있다. <그림 2-6>은 예측의 시·공간을 통해 IPTT와 EPTT의 영역을 보여주고 있다.

IPTT는 초기 Ran의 동적 통행배정 연구에서 예측된 $t+1$, 즉 실제 시간이 속한 시스템 시간대의 구간통행시간 제공정보를 이용하는 차량이 혼잡구간을 즉시 우회한다는 경로전환의 목적으로 도입되었으며(Ran et al., 1993), $t+1$ 의 소통상태가 $t+1$ 이후에도 동일하게 지속된다고 가정한다. IPTT는 일반 ITS시스템에서 구간통행시간 정보를 정보제공 도로단위 구간인 Con-zone 단위로 집계한 후, Con-zone 단위의 구간통행시간 정보를 제공하기 위해 널리 이용되는 개념으로서 엄밀히 말해 장래 통행시간 개념이 아닌 실시간 통행시간의 개념이다. IPTT는 식(2-5)와 같이 정의할 수 있으며, n 은 $o \rightarrow d$ 간 경로에 존재하는 공간적으로 연속된 도로구간의 집합, $x_i(t+1)$ 는 $t+1$ 시점에서 $o \rightarrow d$ 간 경로에 존재하는 i ($i \in n$)번째 구간통행시간, $IPTT_{o,d}(t+1)$ 는 $t+1$ 시점에서 $o \rightarrow d$ 간에 존재하는 도로구간 집합을 주행하는데 소요되는 경로통행시간이다.

$$IPTT_{o,d}(t+1) = \sum_{i=1}^n x_i(t+1) \quad \text{식(2-5)}$$



<그림 2-6> 구간기반 경로통행시간 예측모형의 시·공간

EPTT는 예측모형의 예측영역인 F 가 식(2-4)의 조건을 충족하지 못하는 문제에 대한 대안으로 적용되어 왔다. 즉, 예측된 $t+f$ 의 통행상태가 장래에도 지속된다는 비현실적인 가정에 대한 대안으로서 $t+f$ 이후의 소통상태는 과거의 반복적이고 경험적인 통행시간을 따른다고 가정한다(Peeta and Mahmassani, 1995). <그림 2-6>은 EPTT의 예측영역과 경험영역을 보여주고 있으며, $t+f$ 이후의 구간통행시간은 경험영역인 과거 요일별 시간대별 평균 구간통행시간으로 대체된다.

EPTT는 식(2-6)과 같이 정의할 수 있으며, n 은 $o \rightarrow d$ 간 경로에 존재하는 공간적으로 연속된 도로구간의 집합, t_{sum}^i 은 $o \rightarrow d$ 간 경로에 존재하는 $i(i \in n)$ 번째 구간의 진입시점으로서 $i-1$ 번째 구간의 진출시점까지의 누적 구간통행시간, $x_i(t_{sum}^i)$ 은 i 번째 구간의 진입 시점(t_{sum}^i)에서 구간통행시간, $EPTT_{o,d}(t+1)$ 는 $t+1$ 시점에서 $o \rightarrow d$ 간에 존재하는 도로구간 집합을 주행하는데 소요되는 경험경로통행시간이다.

$$EPTT_{o,d}(t+1) = \sum_{i=1}^n x_i(t_{sum}^i) \quad \text{식(2-6)}$$

IPTT는 실시간 정보수집을 전제로 하지만, EPTT는 ITS시스템의 정보수집 단말이 설치되지 않은 도로구간에 대해서도 경로통행시간 정보를 생성할 수 있는 장점이 있다. 따라서 EPTT는 IPTT의 실시간 구간소통 정보제공과 달리 네비게이션 등으로 제공되는 경로 통행시간의 산정에 주로 적용된다.

5) 구간기반 경로통행시간 예측모형의 문제점

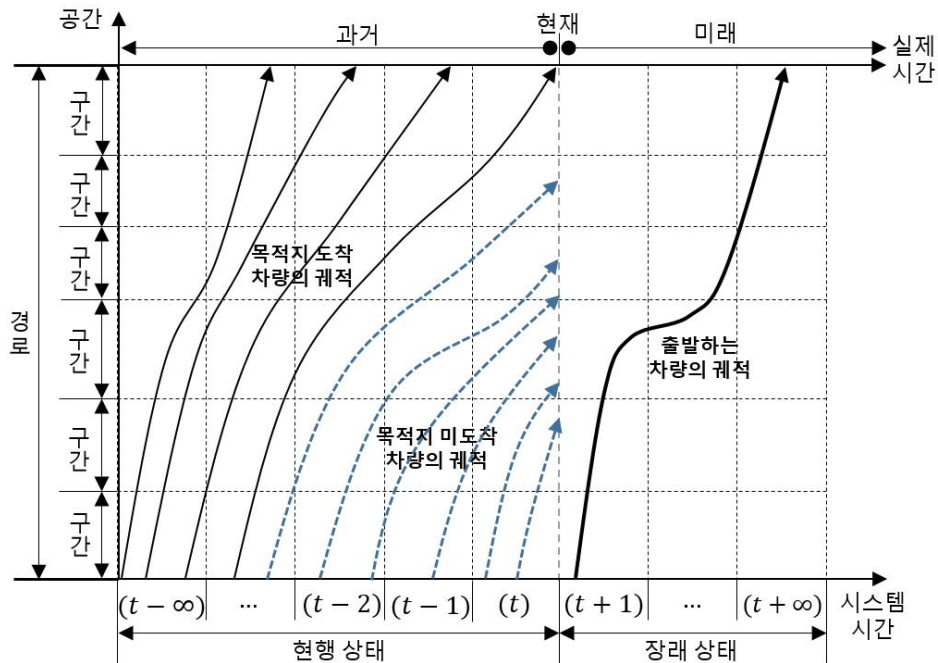
구간기반 경로통행시간 예측모형인 IPTT와 EPTT는 각각 다음과 같은 근본적인 문제점을 가지고 있다. IPTT의 경우 $t+1$ 의 상태가 장래에도 지속된다는 가정을 기반으로 한다. IPTT의 가정이 성립하기 위해서는 교통상태가 정적인(Stationary) 상태진화의 특성을 보여야 한다. 그러나 Vlahogianni et al.(2006)이 밝힌바와 같이 교통상태의 시간적 진화특성 중 정적특성은 비혼잡상태의 일부분을 설명할 뿐 대부분의 상태진화를 설명하지 못하며, 타 연구에서도 이를 뒷받침하고 있다(Altman, 1992; Chang et al., 2012b). 추가적으로 IPTT가 물리학적으로 설명되기 위해서는 $t+1$ 에서 TL 동안에 출발지에서 도착지간의 통행을 완료해야 한다. 그러나 실제로 5분 이내의 자동차 통행비율은 매우 낮다.

EPTT의 경험영역은 주기성(Periodicity)에 근거한 구간통행시간의 반복적(Recurrent) 행태를 가정하며, 주로 요일별 시간대별 링크 통행시간의 평균값을 대푯값으로 이용한다. 그러나 Stathopoulos and Karlaftis(2001)가 밝힌 바와 같이 교통상태는 보편적 개념인 시간과 공간의 주기성(연, 월, 일, 시간)이라는 일반적 개념에서 벗어나며 무작위성(Randomness)을 수반한다. 이는 교통상태 예측을 도전 문제로 만드는 요소이기도 하다(Yoon and Chang, 2014). 추가적으로 주기성이 뚜렷히 나타난다면 주기성을 이용하는 SARIMA의 예측 정확도는 매우 높아야 한다.

4. 경로기반 통행시간 예측모형

1) 경로기반 통행시간 예측모형의 정의

경로기반 경로통행시간(Path-based path travel time, 이하 PPTT) 예측 방법론은 구간통행시간 예측기법의 한정된 예측영역(F)의 확장 한계를 극복하고 LPTT를 대체하기 위하여 Chang et al.(2010)에 의해 제시되었다. PPTT 모형은 다중시간대 구간통행시간 예측의 복잡한 과정을 생략하고 예측 시점에서 출발지 기준의 경로통행시간을 직접 예측하며, <그림 2-7>은 PPTT의 시·공간 예측영역을 보여주고 있다.



<그림 2-7> 경로기반 통행시간의 시·공간 예측 영역

PPTT모형은 장래의 혼잡발생 등으로 인하여 증분(Incremental)된 통행시간을 기존 예측모형의 예측영역(F)을 확장함에 따라 증가하는 불확실성의 문제를 해결할 수 있는 효과적인 방안이다(Chang et al. 2010;

2011). 또한 LTT 예측모형의 현행 상태벡터의 관측값 개수(l)와 예측영역의 개수(f)에 대한 제한이 없다. 즉, l 과 $f \rightarrow \infty$ 으로서 LTT 예측모형의 F 를 무한히 확장하는 것과 같다.

2) 경로기반 통행시간 예측 연구

PPTT 예측모형의 연구에 관한 사례는 많지 않은 실정이며, 고속도로의 IC간 또는 노선버스의 버스 정류장간 PTT를 이용한 연구가 보고되고 있다. 이는 출발지와 도착지간 실제 PTT이 수집이 가능하고 수집자료의 신뢰도가 타 교통정보 수집체계에 비하여 높기 때문이다.

노선버스 PPTT 예측연구의 경우, Chang et al.(2010)은 GPS를 기반으로 하는 자동차량위치(Automatic Vehicle Location)를 통해 수집된 노선버스의 주요 정류소간 PTT와 KNN-NPR 모형을 이용하여 출발지에서 도착지까지 PPTT를 예측하였으며, 시간적으로 45~90분의 PTT 정보수집에 요구되는 시간 처짐(Time lag)을 극복하면서 PTT를 예측할 수 있음을 보였다. 후속 연구로 An et al.(2013)은 노선버스의 이력자료와 결합된 확률과정 모형을 이용하여 버스경로통행시간을 예측하였으며, 모형의 정확도는 KNN, 확률과정, 이력 평균의 순으로 나타나 KNN모형의 예측성능이 보다 우수함을 보였다.

고속도로 PPTT 예측연구의 경우, 남궁성(2005)은 한국도로공사의 요금징수시스템인 TCS(Toll Collection System)에서 수집되는 인접 TG간 PTT를 이용하여 고속도로 TG간 PTT를 추정하는 전진반복전후방탐색(PIFAB) 모형을 개발하고, 이를 이용하여 고속도로 TG간 중·장거리 PTT를 추정하였다. Chang et al.(2011)은 TCS 경로통행시간을 이용하여 출발시각 기준의 중·장거리 PTT를 KNN-NPR 모형을 이용하여 예측하였으며, PTT의 수집에 요구되는 시간처짐을 효과적으로 극복할 수 있었다. 모형의 입력값으로 톨게이트 유·출입 통행량을 이용하여 고속도로 망에서 변화하는 교통류 상태를 고려하고자 하였으며, 교통상태의

예측에 있어 도로망의 유·출입 통행수요 만으로도 PTT의 예측이 가능함을 보였다. 또한 도로망의 유·출입 수요는 도로망 전체의 소통상태를 반영하는 주요 변수임을 지적하였다. 이후 Bustillos and Chiu(2011)는 유·출입 교통량만을 이용한 KNN모형을 이용하여 고속도로 PTT를 예측하였다.

PPTT 모형은 노선버스, 고속도로 등의 고정 노선을 경유하는 통행이 높은 환경에서 효과적인 방법이다. 반면, PPTT 모형은 충분한 과거 PTT Database와 현행 상태의 정보를 이용하여 신뢰성 높은 과거의 PTT 정보를 추출할 수 있는 신뢰성 있는 정보가 요구된다. 따라서 향후 대용량 기반의 교통 소통자료의 수집 및 관리가 현실화될 경우, 그 적용성은 증가할 것으로 판단된다.

3) 경로기반 통행시간 예측의 문제점

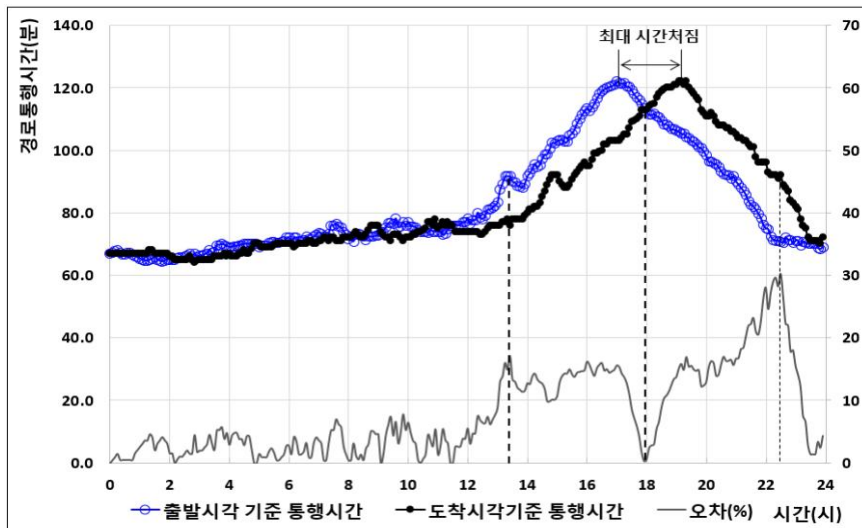
PPTT 모형은 LPTT 모형보다 정확한 경로정보 예측이라는 점에서 효과적인 방법론이며(Chang et al., 2010), Chen과 Chein(2000)은 LPTT 평가 방법론에 비하여 PPTT 평가 방법론이 우수함을 보였다. 이러한 PPTT 모형은 예측의 본질적인 문제인 불확실성을 감소시킬 수 있는 하나의 대안일 수 있다.

반면, PPTT 모형은 다양한 패턴이 내재되어 있는 방대한 양의 과거자료가 요구되며, 구축된 자료의 질도 우수해야 한다는 단점이 있다. 또한 천문학적인 기·종점간 PTT 자료를 구축할 수 없기 때문에 일반적인 경우의 PTT 예측에 적용할 수 없다는 한계가 있다. Chang et al.(2010)의 경우, 버스의 역별 승하차 인원, 동일 버스노선의 실시간 통행정보 등을 고려하지 않았다. 또한 Chang et al.(2011)의 경우, 고속도로 구간의 교통상황을 고려할 수 없는 단점이 있었다. 그러나 방대한 양의 과거자료를 이용할 수 있는 대용량 데이터베이스 환경이 조성되면 효과적인 경로통행시간 예측기법의 하나라고 판단된다.

5. 기존 통행시간 모형의 한계 및 문제점

1) 도착시각 기준 통행시간 모형

도착시각 기준 경로통행시간(Arrival time based path travel time, 이하 APTT)은 프로브의 경로통행이 완료된 후 수집된 개별 PTT를 출발지의 출발시각으로 집계한 PTT로서 현 시점에서 APTT 만큼의 이전에 출발한 차량의 PTT를 의미한다. 따라서 실제 출발시각 기준 경로통행시간(Departure time based path travel time, 이하 DPTT)의 통행시간에 해당하는 시간차집(τ)이 발생한다. <그림 2-8>은 APTT와 DPTT의 시간적 차이를 시계열자료로 보여주고 있다. 비혼잡 상태는 유사한 결과를 보이지만 τ 을 수반하고 있다. 그리고 혼잡이 발생하게 되면 혼잡으로 인한 통행시간의 증분에 비례하여 APTT의 τ 는 증가하며, 통행시간 증가 시 τ 의 증가는 오차의 증가로 이어진다. 6~24시의 평균 오차는 9.88%로 분석된 반면, 혼잡시간대인 13~24시의 오차는 최대 30.2%까지 증가하면서 평균 14.3%로 높게 나타났다. 따라서 APTT의 경우 허용오차 $\pm 10\%$ 내에서 중·장거리 통행시간의 정보제공에는 부적합하다고 판단된다.

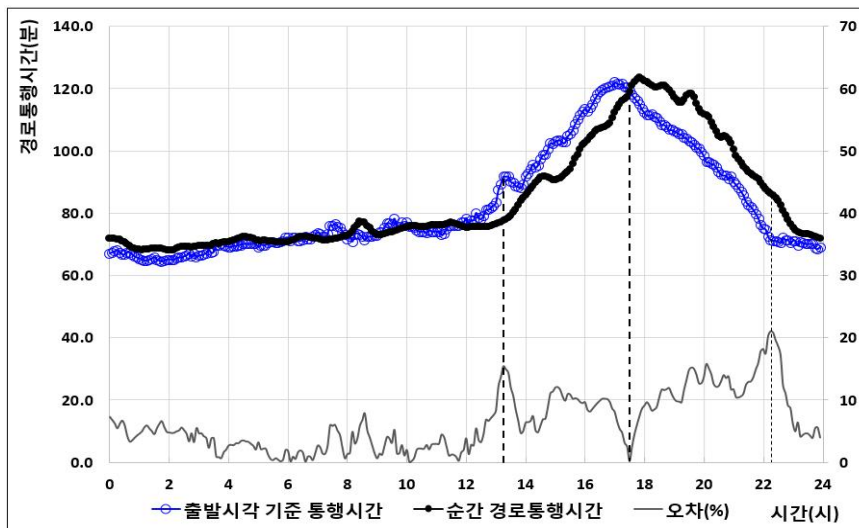


<그림 2-8> 출발시각과 도착시각 기준 경로통행시간의 비교

2) 구간기반 경로통행시간 모형

<그림 2-9>는 IPTT와 DPTT를 시계열로 보여주고 있으며, IPTT는 본 연구의 자료구축에 이용된 구간거리(0.6~13.0km, 평균 4.5km)의 13.0km의 구간통행시간 자료의 수집에 요구되는 τ 을 고려한 $t-\tau$ 의 시점에서 $t+1$ 의 구간통행시간을 100% 정확히 예측했다는 가정하에 산정하였다. 따라서 최장거리 구간인 13.0km를 혼잡시 통행속도 10km/시로 주행했다고 가정하고 $TL=5$ 분으로 설정하면 $\tau \approx 16$ [$15.6 = (60.0 \times 13/10)/TL$]이며, $t+1$ 의 장래상태 예측을 고려하면 실제 자료수집 시점에서 예측영역(F)은 17까지 연장되며, 83분 후의 구간통행시간을 정확히 예측한 것과 동일하다.

비 혼잡시 IPTT와 APTT는 유사한 결과를 보이고 있다. 그리고 혼잡발생시, IPTT는 APTT의 시간처짐을 상당부분 극복한 것으로 판단되나, 통행시간의 증가/감소시 상태회복(State recovery)³⁾에 실패하고 있다. 따라서 6~24시의 평균 오차는 7.32%인 반면, 혼잡시간대의 오차는 최대 21.1%까지 증가하면서 평균 10.6%로 나타났다. 허용오차 $\pm 10\%$ 내에서 중·장거리 통행시간의 정보제공에는 적합하지 않다고 판단된다.



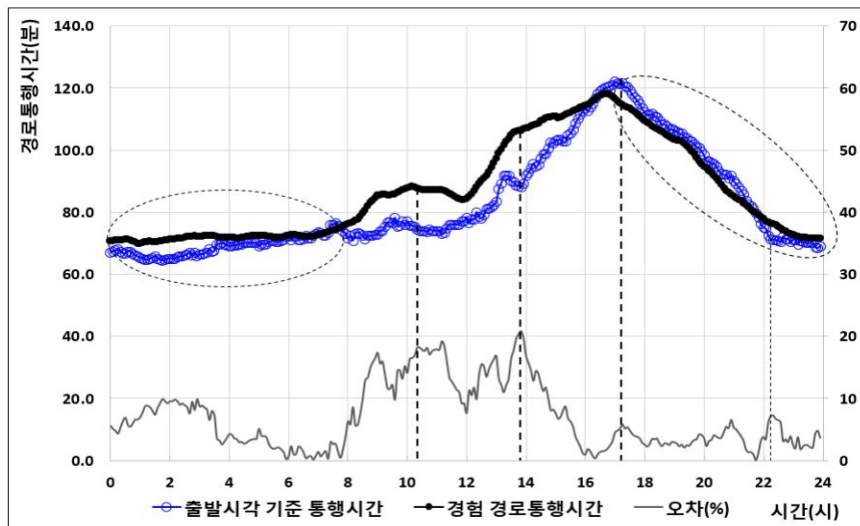
<그림 2-9> 출발시각 기준과 동시 경로통행시간의 비교

- 3) 본 연구에서 상태회복은 실제 교통상황 변화시점 이후의 변화상태를 모형이 인지하고 변화된 상태의 방향성과 변화량을 재구성함을 의미한다.

3) 경험기반 경로통행시간 모형

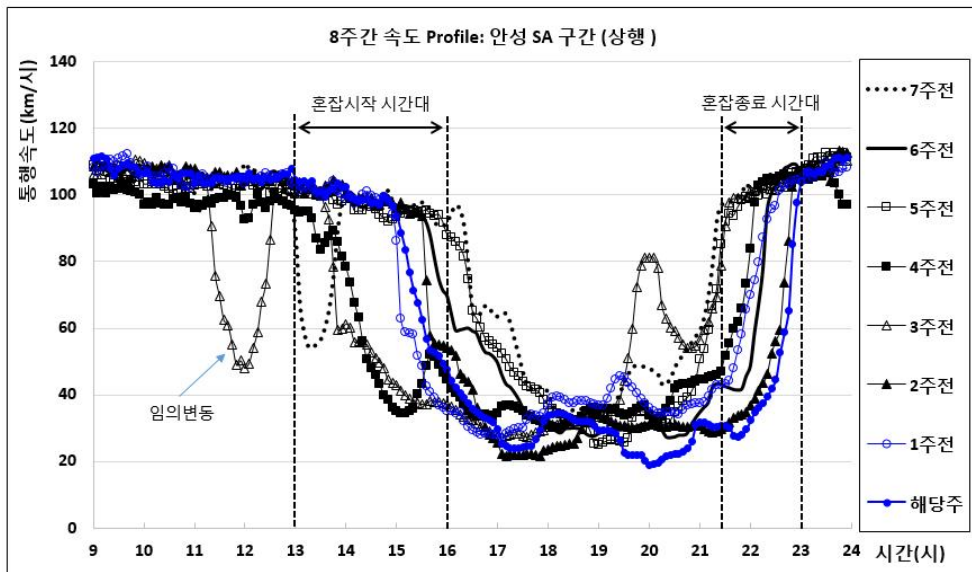
EPTT는 예측범위 이후의 통행시간을 과거 통행시간 Profile로 대체한 후 산정하였으며, <그림 2-10>과 같다. 그리고 예측범위와 예측 정확도의 가정은 <그림 2-9>의 IPTT와 동일하다. EPTT의 경우, 자유교통류 상태가 지속되는 반복적 행태를 보이는 심야에서 출근 시간대 이전까지 그리고 (중·장거리 통행의 시점에서) 고속도로 교통량이 소거되지 시작하는 18시 이후의 준 반복적 행태에서 10% 내외의 예측오차를 보이고 있다.

평균오차는 6~24시 7.30%, 혼잡시간대 6.21%로서 APTT와 IPTT보다 우수한 것으로 나타났다. 그러나 혼잡 발생 및 성장 시간대의 최대 오차는 20.9%까지 증가하고 있으며 8~16시의 평균 오차는 13.33%로 나타났다. 따라서 EPTT의 경우, 허용오차 $\pm 10\%$ 내에서 주간 시간대 중·장거리 통행시간의 정보제공에는 적합하지 않다고 판단된다. 이는 중·장거리 통행의 경우, 통행의 상당부분이 지역간 고속도로로 구성되며, 지역간 고속도로 구간은 도시부 고속도로 구간에 비하여 구간통행시간의 반복적 특성이 낮고 통행 수요의 임의성이 높기 때문이다.



<그림 2-10> 출발시각 기준과 경험 경로통행시간의 비교

지역간 고속도로 구간의 혼잡은 어느 정도 주기성을 보이면서 임의성(Randomness)과 변동성(Variability)이 크게 나타나는 특성이 있다. <그림 2-11>은 2010년 11월 19일(금요일 기준)부터 과거 7주까지 경부선 안성 휴게소 구간의 통행속도를 시계열로 보여주고 있으며, 혼잡이 시작되는 시간대는 13~16시까지 3시간 내에서 그리고 소거 시간대는 21시 30분~23시까지 1시간 30분 내에서 변화하고 있다. 혼잡 지속 시간대는 4.5~9시간까지 넓게 형성되고 있으며, 혼잡의 성숙단계에서 속도는 20~40kph로 다르게 형성되고 있다. 이는 최소한 IPTT 추정이 가능한 환경에서 EPTT 모형을 이용하여 경로통행시간 정보를 제공할 경우, 통행시간의 동적특성을 상실하면서 예측오차를 증폭시킬 수 있음을 의미한다. 만일 EPTT가 IPTT보다 예측력이 우수하게 나타난다면 이는 거시적 교통류 시스템의 상태진화 행태가 주기성을 보이는 확률적 상태(Stochastic state)라는 간접적 증거이며, 주기성과 확률적 상태이론을 기반으로 하는 SARIMA와 같은 시계열 분석기법의 예측력이 우수해야 한다.



<그림 2-11> 혼잡의 시간적 주기성과 변동성

4) 인위적 모델링의 한계

일반적으로 예측문제는 교통상태의 진화 행태에 대한 깊은 이해가 선행되어야 한다. 이는 현재의 첨단 예측모형의 대부분이 특정 상태이론(State theory)을 기반으로 개발된 이유이기도 하다. 반면, 현재까지 교통류 상태의 시·공간적 진화과정에 대한 행태와 이론은 명확히 밝혀지지 않은 상태이다. 이는 혼잡계(Complex system)로 대표되는 대다수 상태 시스템에 대한 공통된 도전과제라 할 수 있다.

따라서 ITS가 도입된 이후 교통상태를 예측하기 위한 모형의 개발자들은 교통류 행태의 경계상태(Boundary condition), 즉 상태의 전환점(Turning point)을 사전에 추정하기 위하여 많은 노력을 기울여 왔다. 특정 상태이론을 기반으로 교통상태 예측모형을 개발하는 대다수의 모델러들은 대상 시스템의 상태내에 내재된 설명할 수 없는 행태와 알려지지 않은 변수(Unknown parameter)에 대하여 상당 부분을 인위적 판단에 의존하게 된다. 이는 인간이 소유한 상태이론과 그 상태를 설명하기 위한 지식의 한계이기도 하다. 이를 극복하기 위하여 신경망과 같은 비모수 계열의 예측모형들이 제시되고 있지만, 통계기반의 모형보다 다소 우수한 결과를 도출하고 있는 정도이다. 폭넓은 기존연구의 검토결과, 모형 개발자의 인위적 모델링이 교통류 시스템의 상태진화를 얼마나 설명할 수 있는가?라는 본질적 의문을 갖게 한다.

5) 적용 및 유지·보수의 한계

21세기의 고도화된 예측모형은 실제 ITS 시스템에 탑재된 후 현장의 운영요원이 운영·관리에 있어 극복해야할 새로운 문제를 낳고 있으나, 모형 개발자는 이를 고려하지 않고 있다(Smith and Oswald, 2003; Chang et al., 2012b). 고도로 정교한 모형을 이해하고 이를 시스템에서 운영하기 위해서는 탑재된 예측모형이 기반한 특정 상태이론과 더불어 교통류 이론에 대한 폭넓은 지식과 이해, 그리고 모델링에 대한 폭넓은

경험이 필요하다(Yoon and Chang, 2014). 그러나 Smith and Oswald(2003)가 지적한 바와 같이, 현장의 유지·관리 요원은 많은 경우에 있어 운영 시스템과 데이터베이스관련 풍부한 경험과 지식을 소유한 반면, 고도화된 예측모형에 대한 이해와 지식은 부족하다. 따라서 ITS 시스템을 유지·관리하는 현장의 요원들은 고도화된 예측모형의 모델링, 모형의 구조변경 그리고 입력 파라미터의 지속적인 재정산 등에 대한 문제를 경험하고 있다.

6. 본 연구의 차별성과 기여도

1) 연구의 차별성

기존의 교통상태 예측모형은 단일 (또는 몇 개의 인접) 지점 또는 구간에 대한 소량의 정보를 이용하여 단기예측(Short-term prediction)의 예측영역으로 범위를 한정하여 예측의 정확도 향상을 위한 모형의 개발과 고도화를 수행해왔다. 이러한 시간적으로 한정되고 모델링 측면에서 고도화된 교통상태 예측연구는 실시간 정보제공 및 교통제어의 측면에서 ITS의 편익을 증대시켜왔다. 반면, 통행시간 정보제공의 측면에서 일반적으로 5~15분 이내인 단기 예측모형의 예측영역 이상인 통행시간에 대하여 장래의 소통상태를 고려한 합리적인 통행소요시간 정보를 제공할 수 없다는 한계를 가지고 있다.

본 연구는 기존 단기예측모형의 지속적인 도전 목표였던 수용가능한 예측오차의 범위내에서 장래 6시간까지 예측영역의 확장이 가능한 교통소통상태 예측모형을 개발하고, 예측된 소통상태 즉 구간통행속도를 이용하여 고속도로가 경유하는 도시의 IC/TG간 중·장거리 경로통행시간 정보의 예측에 중점을 둔다. 이를 위해 본 연구에서 모형의 개발과 적용을 위해 고려한 사항은 다음과 같다.

첫째, 기존의 일반적인 교통상태 예측모형의 개발 방향이었던 특정 상태 이론을 기반으로하는 인위적 모형개발 방식에서 벗어나 방대한 양의

데이터베이스에 내재된 정보를 이용하는 모형을 개발하였다. 데이터베이스로부터 장래 교통상태에 대한 지식을 추출하기 위하여 개발모형은 지식발견(Knowledge discovery) 접근법을 기반으로 개발되었다. 다시말해 특정 상태이론을 기반으로 교통상태의 진화과정을 모델링함으로써 장래 상태를 추정하는 기존의 접근법과 달리 본 연구에서 개발된 모형은 탐색과 지식발견의 과정을 통해 추출된 교통류 상태에 내재된 진화과정의 상태변화를 이용하여 교통상태를 예측한다.

둘째, 교통혼잡의 공간적 진화행태는 독립된 구간으로 설명할 수 없는 한계가 있다. 따라서 본 연구에서는 기존 교통소통상태 예측모형의 공간적 대상인 단일 지점 또는 구간의 공간적 규모를 구간의 집합인 도로망 차원으로 확장함으로써 도로망 전체의 소통상태, 즉 혼잡의 발생·진화·성숙·완화·소멸의 상태를 도로망 수준에서 설명할 수 있도록 하였다. 따라서 구간통행시간의 직접적 예측이라는 기존방식에서 벗어나 도로망 소통상태를 예측하고, 예측된 도로망 소통상태를 이용하여 경로통행시간을 추정함으로써 장래 불확실성에 대한 영향을 최소화하고자 하였다.

셋째, 교통류 시스템의 시간적 상태진화는 구간으로 연결된 도로망에서 형성되며, 교통혼잡의 시간적 발달과정은 매우 가변적이며 다양한 형태로 나타난다. 따라서 본 연구에서는 혼잡의 시간적 진화과정을 단일 구간의 시계열적 상태변화가 아닌 도로망의 차원에서 시간에 따른 혼잡의 생성과 소멸을 설명하도록 한다. 시·공간적으로 확대된 네트워크 기반 접근법은 소통상태의 예측에 있어 불확실성을 감소시킬 수 있는 방안으로서 기존 단기예측모형의 시·공간적 한계를 극복하고자 하였다.

넷째, 기존 단기예측기법의 한계였던 5~15분의 예측영역을 수용가능한 예측오차의 범위내에서 장래 6시간까지 확장하였다. 이는 물리학적 시간영역을 위배하지 않는 상태에서 중·장거리 통행시간의 제공이 가능함을 의미한다. 장래 소통상태를 고려한 주요 도시간의 통행시간정보를 제공할 수 있음을 의미한다.

다섯째, 모형의 개발시 실제 첨단자료관리시스템의 탑재를 목표로

Off-line 모형개발이 아닌 On-line 적용을 고려하여 모형을 개발하였다. 이를 위해 데이터베이스에 포함된 누락자료, 현행 자료의 수집에 요구되는 시간차집 등 자료환경을 고려하여 모형을 개발하였다.

여섯째, 다양한 ITS시스템의 적용을 고려하여 모형의 연산속도를 극대화하도록 모형을 설계함과 더불어 모형의 파라미터를 최소화하였다. 파라미터 값은 모형의 연산과정에서 자동으로 최적화되도록 하였다. 그리고 입·출력 정보를 용이하게 구조화할 수 있도록 모형의 구조를 설계함으로써 개발모형의 적용성을 극대화하고자 하였다.

마지막으로 개발모형의 시스템 탑재 후 지속적인 유지·관리를 고려하여 교통류와 모델링에 대한 폭넓은 지식이 아닌 데이터베이스와 시스템에 대한 경험과 지식이 풍부한 시스템 유지·관리 요원의 입장에서 운영이 가능하도록 데이터기반의 모형을 개발하였다.

2) 연구의 기여도

앞에서 설명한 연구의 차별성과 더불어 본 연구의 시대적 측면, 학문적 측면, 기술적 측면, 그리고 실무적 측면에서 기여도는 다음과 같다.

첫째, 시대적 측면으로 데이터가 변하고 있다. 기존의 지점검지체계는 개별 차량기반의 프로브 방식으로 확장되었으며, 그 범위는 전국의 모든 도로를 대상으로 하고 있다. 다시말해 데이터의 양과 질, 시·공간적 커버리지가 전국의 모든 도로로 확대되면서 대용량자료의 시대에 진입하였다. 이는 실시간 적용을 목표로 하던 기존의 지점 또는 구간기반의 단기 예측모형이 다른 형태의 예측모형으로 변해야 함을 의미한다. 그 변화의 방향중 하나는 본 연구에서 시도하는 도로망 차원에서 단·중·장기 소통상태를 예측하는 것이다. 이를 통해 보다 먼 미래의 도로망 소통상태 정보를 제공할 뿐만 아니라 이론적으로 합리적인 중·장거리 통행시간정보를 제공함으로써 ITS의 편익을 증대시킬 수 있을 것이다.

둘째, 학문적 측면에서 기존 단기에측의 예측영역 확장이라는 벽을 허

물수 있는 기회이다. 국내에 ITS가 도입된 이후 도시·교통분야의 예측 분야는 하나의 학문으로 정착되었다. 그러나 현재까지 많은 전문가의 노력에도 불구하고 제2의 부흥기에 접어들지 못하고 있다. 실례로 단속류의 변동이 심한 교통류상태를 허용오차 10%이내에서 예측하지 못하고 있는 수준이며, 철저한 기존연구 고찰결과 극심한 교통류의 변동상태에서 단일 시간대에 있어 10%이내에 진입한 연구(Vlahogianni et al., 2005; Yoon and Chang, 2014)는 극소수에 불과한 실정이다. 일부 연구에서 예측영역은 1시간까지 확장되었지만 특정 도로구간 또는 교통수단에 한정하되 있다. 수용가능한 예측오차의 범위내에서 도로망 차원의 중·장기 통행시간 예측은 아직까지 보고되고 있지 않다. 따라서 본 연구와 같은 대용량 데이터기반의 예측연구는 향후 ITS 예측분야 연구에 있어 새로운 문제해결 방식과 이론을 형성하는 계기가 될것으로 판단된다.

셋째, 기술적 측면으로 ITS Hardware 선진국에서 정보활용 기술의 선진국으로 변해야 한다. 국내의 경우 ITS분야는 이미 대용량자료의 시대에 들어섰지만, 대용량 자료에 적합한 예측기술은 제시되지 않고 있는 실정이다. 선진국의 경우 기존 ITS체계하에서 수집되는 정보를 이용하여 다양한 형태의 장래 상태 예측기를 개발함으로써 장래 소통정보를 제공하고 있다. 독일은 미시적 모의실험기를 기반으로하는 OLSIM을 개발하여 독일 아우토반 고속도로의 장래 소통정보를 온라인으로 제공하고 있으며 (<http://www.autobahn.nrw.de>), 프랑스는 현행자료와 이력자료를 조합한 통계기반의 소통상태 예측모형인 ATHENA를 개발하여 활용하고 있다 (<http://www.inrets.fr/linstitut/unites-de-recherche-unites-de-service/grettia/>). 국내의 경우, 도로분야 ITS에서 한국도로공사와 같은 선두적 역할을 수행하는 관련기관은 이에 대한 연구를 추진하고 있는 반면, 다양한 형태로 분포해 있는 지자체 및 타 기관 ITS시스템의 실질적인 내부구조는 기존의 ITS와 큰 차이를 보이고 있지 않다. 따라서 본 연구에서 개발되는 KJC 예측기는 기존의 20세기 ITS 시스템에서 내부구조가 변화된 21세기의 ITS로 변화를 시도하고자 하는 일련의 노력이라 할 수 있다.

마지막으로, 실무적 측면에서 이력자료의 활용방안을 제시하고 있다. 우리나라는 Hardware측면에서 ITS 선진국이라 해도 과언은 아니다. 그러나 국내의 대다수 ITS시스템의 경우 저장·관리되는 방대한 양의 이력자료에 대한 활용방안을 찾지 못하고 있는 실정이다. 엄밀히 말하자면, 동적 교통정보의 생성 및 가공에도 단기 예측모형조차 이용하지 않고 있는 경우가 많으며, 수집·저장된 방대한 자료를 이용할 수 있는 소규모 자료지원시스템조차 하부시스템으로 구성되지 않은 경우가 많다. 본 연구는 ITS시스템의 일차적 목표인 교통정보의 수집·제공 차원을 보다면 장래의 정보로 확장함으로서 시스템의 DB공간 안에서 잠자고 있는 이력자료의 활용방안을 위한 핵심기술을 제공하고 있으며, 이를 통해 ITS의 편익을 증대할 수 있을 것이다.

Ⅲ. 대용량 자료기반 통행시간 예측모형 개발

1. 모형의 개발과정 정립

1) 개발방향의 설정

장래의 교통상태를 100% 완벽하게 재현하는 교통 시뮬레이터가 있다고 가정해 보자. 그렇다면 예측모형을 이용하여 통행시간 정보를 생성할 것인가? 아니면 시뮬레이터의 결과를 이용할 것인가? 아마도 많은 전문가들은 후자를 선택할 것이며, 통행시간 예측의 방법론도 지금과는 매우 다를 것이다. 그렇다면 완벽한 교통 시뮬레이션 결과가 대용량 교통 데이터베이스 안에 있는데 그걸 이용하지 않는가?

서론에서 언급한 바와 같이 현대의 교통정보 수집원, 그리고 자료의 질과 양은 기존의 단일 ITS 시스템에서 수집/관리하던 정보와는 매우 다르다. Smart phone의 네비게이션 어플과 DSRC 단말기를 통해 수집되는 교통소통정보는 IT의 기술 발달과 더불어 높은 정확도를 보유하고 있으며, 수집되는 정보의 양과 공간적 규모는 국내에서 가장 큰 대용량 데이터라고 표현할 수밖에 없다. 그러므로 가용자료의 양과 정확도가 급변하는 시점에서 교통상태의 예측방법론도 그에 부합하게 변화해야 한다.

본 연구에서는 기존 통행시간 예측기법의 한계를 극복함으로써 보다 먼 장래의 통행시간을 수용 가능한 오차의 범위 내에서 예측하고, 첨단 자료관리시스템에 탑재가 용이하며, 교통류의 행태와 예측 모델링에 대한 전문지식이 없이도 운영이 가능한 데이터 기반의 예측방법론을 개발하고자 한다. 이러한 예측모형의 개발방향은 모형에 데이터를 맞추던 기존의 방식을 벗어나, 모형이 데이터베이스를 탐색함으로써 선별적으로 데이터를 받아들이고 선별된 데이터를 기반으로 하는 지식발견(Knowledge discovery, 이하 KD)과 의사결정의 과정을 통하여 장래 소통상태를 예측하는 지식 탐색형 모형개발을 의미한다.

2) 요구조건 분석

예측모형의 설계와 개발에 앞서 모형의 용도와 운영환경 등을 고려한 요구조건이 선행되어야 한다. 특히, 본 연구에서 개발하고자 하는 예측모형은 첨단자료관리시스템(ADMS)과 같은 자료관리시스템의 방대한 데이터베이스를 이용하여 단거리뿐만 아니라 중·장거리 통행시간의 예측을 목표로 한다. 따라서 데이터베이스의 요구조건, 파라미터의 정산, 예측모형 전문가가 아닌 일반 현장 요원의 운영 편리성 등의 다음과 같은 요구조건을 고려하여 모형을 개발하도록 한다.

첫째, 예측영역 확장에 따른 불확실성(Uncertainty)의 극복방안이다. 임의성을 수반하는 교통상태의 동적 특성을 고려하면 예측상태에 대한 불확실성은 예측/추정 기법의 근본적인 문제라 할 수 있다. 단·중·장거리 경로통행시간을 예측하기 위해서는 기존의 단기 구간통행시간 모형의 예측시간 영역인 5~15분을 최소한 6시간까지 확장해야 한다. 따라서 도로망 수준에서 도로망의 전반적인 시간대별 소통상태를 6시간까지 예측하고, 예측된 도로망의 하위 구성집합인 구간별 소통상태를 기반으로 장래의 소통상태 변화를 고려한 경로통행시간 예측모형을 개발하도록 한다.

둘째, 개발모형의 공간적 이식성 즉, 다양한 ITS시스템의 탑재이다. 일반적으로 예측모형 모델러는 특정 상태이론에 근거하여 예측모형을 개발하게 되며, 개발과정에서 대상 시스템의 상태 진화, 내재된 행태 그리고 알려지지 않은 파라미터 등에 대한 모델러의 인위적 판단이 추가된다. 따라서 개발된 모형을 타 시스템에 적용할 경우 모형의 적용성과 예측력은 낮아지게 된다. 반면 본 연구에서는 모델러의 교통류 이론에 대한 배경지식이나 특정 상태이론을 가정하지 않고 개발모형의 탑재 대상인 ADMS에서 수집·관리하는 방대한 양의 데이터베이스에 의존하는 탐색기반의 지식발견 접근법을 기반으로 모형을 개발하도록 한다. 따라서 본 연구에서 개발되는 모형은 운영중인 ITS시스템이 수집하는 해당 자료의

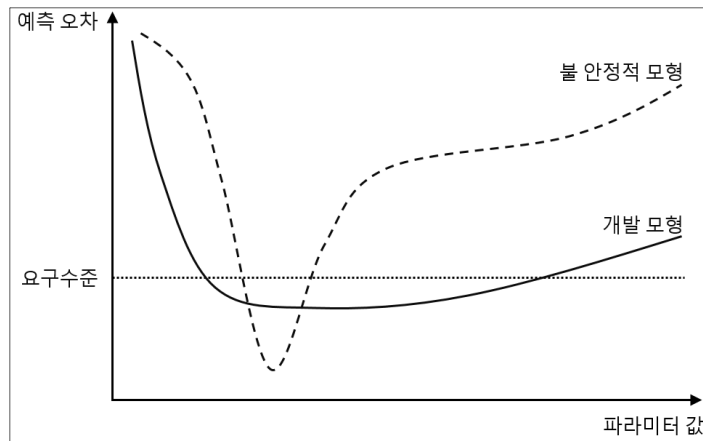
상태이론을 고려할 필요가 없으며, 전적으로 해당 시스템이 수집·관리하는 DB에 의존하기 때문에 모형의 공간적 이식성을 유지할 수 있다.

셋째, 시스템이 보유한 다양한 데이터의 활용이다. 교통상태의 진화특성은 교통상황에 따라서 매우 동적이며 다양하게 나타난다. 교통자료(교통량, 속도, 점유율)간의 관계는 다양한 정보가 혼재되어 있으므로 결론적으로 규정하기는 참으로 어려운 일이다(Disbro and Frame, 1989; Kerner, 1996; Vlahogianni et al., 2006; Chang et al., 2007). 그러나 과거 이력자료의 다양한 변동에는 많은 정보를 포함하고 있다. 따라서 모형의 개발시 이력자료에 포함되어 있는 다양한 정보를 이용할 수 있는 다중입력체계를 기반으로 모형을 개발하도록 한다.

넷째, 데이터베이스에 내재된 다양한 상태정보의 특성이다. 확률적 상태이론(Stochastic state theory)에 의하면 교통상태의 복잡한 내재적 특성은 주중/주말, 각 요일별 시간대별 패턴과 같은 주기성(Periodicity)과 시계열 변동이 내재되어 있다. 반면 혼돈 상태이론(Chaotic state theory)에 의하면 국부적 시계열자료의 국부적 변동은 다양한 장래상태의 정보를 포함하고 있다. 따라서 시계열자료에 평활화 과정을 적용할 경우 시계열자료의 변동과 내재된 정보는 희석되며, 추정되는 상태의 시계열적 변동에 대한 민감도는 낮아지게 된다(Yoon and Chang, 2014). 본 연구에서 개발되는 모형은 데이터의 내재적 특성이 포함되어 있는 원자료를 그대로 이용하도록 하며, 인위적인 시계열적 변동과 특정 시간대별 주기성의 고려가 아닌 모형의 지식탐색 과정에서 내재적 특성을 이용하도록 한다.

다섯째, 데이터베이스에 포함된 누락자료의 처리문제이다. 일반적으로 대용량 시스템은 일정부분 누락자료를 포함하고 있다. 누락자료는 메인 시스템의 일시적 장애, 현장에 설치된 정보수집 장치의 일시적 장애 및 고장, 통신장애 등 다양한 원인으로 발생하게 됨으로 누락자료의 형태는 시·공간적으로 다양하게 나타난다. 따라서 개발모형은 다양한 형태의 누락자료가 존재하는 환경하에서 효과적으로 이를 처리하도록 한다.

여섯째, 모형의 예측력에 대한 지속적인 안정성이다. 모형이 지나치게 정적이거나 민감하지 않은 경우 교통상태의 동적특성에 대한 설명력은 낮다. 반면, 모형이 지나치게 민감한 경우 교통상태의 동적특성을 고려할 수 있으나 최적 파라미터값이 지속적으로 갱신되지 않으면 예측오차는 증폭될 수 있으며, 이는 고도화된 모형의 일반적 특징이다. 따라서 개발되는 모형은 요구되는 예측오차의 범위에서 최적 파라미터 값의 영역을 확대함으로써 안정적인 모형을 개발하도록 한다.



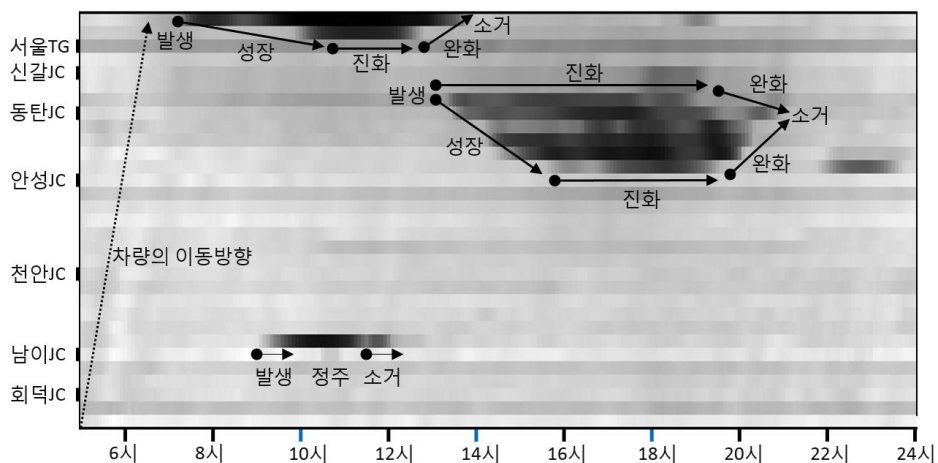
<그림 3-1> 개발모형의 파라미터값에 따른 오차공간 목표

일곱째, 모형의 구조변경 용이성이다. 일반적으로 시스템을 유지·관리하는 현장의 요원들은 데이터베이스에 폭넓은 지식을 보유하고 있지만, 상대적으로 교통류 이론과 고도화된 예측모형에 대한 경험은 적은 편이다. 따라서 현장조건의 변화와 수집자료의 추가 등으로 모형의 입력자료 환경이 변화될 경우 운영중인 모형의 구조변경, 알고리즘 수정, 이에 따른 파라미터의 재정산 등이 필요하다. 추가적으로 이는 해당 사업에 대한 시간과 예산의 문제이기도 하다. 따라서 본 연구에서는 입·출력자료의 구조변경이 용이하도록 모형을 개발하도록 하며, 이에 따른 파라미터의 정산이 간편하도록 모형을 개발한다.

마지막으로 현장에서 운영문제이다. 모형의 정확도가 요구수준을 만족하면 모형의 연산 수행속도와 더불어 운영상의 용이성을 고려해야 한다. 모형의 예측력이 뛰어나더라도 시스템 측면에서 연산 수행시간이 예측정보를 활용할 수 있는 시간적 범위를 초과하거나, 운영측면에서 지속적으로 모형의 파라미터를 조정해야 한다면 실제 모형의 적용성은 현저히 저하된다. 따라서 파라미터의 반복적 정산이 시스템적으로 간편하면서 모형의 연산속도가 빠른 모형을 개발하도록 한다.

3) 개발모형의 개념 및 접근법

<그림 3-2>는 경부고속도로 대전IC→서울TG 상행구간의 시간에 따른 혼잡의 발생-시·공간적 진화-소멸의 과정을 보여주고 있으며, 혼잡의 시·공간적 진화행태는 다양하게 나타나고 있다. 안타깝게도 지점 및 구간에서 교통류 이론은 존재하지만 도로망에서 어느 시각에 어느 정도의 혼잡이 시·공간적으로 어떻게 발생·진화할 것이며 언제 해소될 것인지를 명확하게 설명하는 교통류 이론은 존재하지 않는다. 이러한 명확하지 않은 혼잡의 동적인 시·공간 상태진화 특성은 결국 구간단위의 미시적인 중·장기 소통상태 예측에 있어 불확실성의 문제로 귀착된다.



<그림 3-2> 혼잡의 동적 생애주기 및 다양성

본 연구에서 혼잡을 바라보는 관점은 구간 또는 몇 개의 연속된 구간으로 한정되는 기존의 미시적 관점에서 벗어나 거시적인 도로망 차원으로 확장된다. 그리고 예측모형의 개발은 “교통혼잡은 도로망에 스스로 들어오고, 일정기간 동안 빠져 나가지 못하며, 자신도 모르게 도로망을 탈출한다”는 혼잡의 행태에 대한 개념에서 출발한다. 설정된 혼잡행태의 개념을 확장하여 혼잡의 시·공간적 행태를 설명하자면 다음과 같다. 혼잡은 수요를 따라 폐쇄된 도로망에 진입한 후, 자신의 모습을 드러내기 전에 미세하지만 충분한 신호(Signal)를 노출한다. 그리고 초기 진화과정의 끝에서 모습을 드러낸 이후에는 시·공간적으로 다양한 패턴으로 성장하며, 수요를 따라 도로망을 탈출한다.

이상의 개념을 기반으로 도로망 차원의 소통상태를 예측하기 위하여 본 연구에서는 혼잡의 시·공간적인 잠복, 발생, 성장, 완화, 소멸에 대한 일련의 진화과정에 대한 신호로서 교통류 변수를 이용하되 교통류의 상태 진화에 대한 어떠한 이론도 전제하지 않는다. 반면 상태이론을 대신하여 첨단자료관리시스템의 데이터베이스에 내재된 교통류 상태진화에 대한 방대한 양의 지식(Bulk of knowledge)를 이용하도록 하며, 지식의 추출은 지식발견(KD) 접근법을 이용한다. 예측과정은 교통류 변수의 상태벡터에 내재된 신호와 방대한 양의 시스템 이력자료에 포함된 신호간의 유사성(Similarity)을 탐색하고, 탐색된 유사성을 이용하여 도로망 소통상태를 예측하는 일종의 지식탐색 기반 장애 의사결정과정이다.

4) 모형의 개발 과정 정립

모형의 개발을 위한 개념과 접근법을 기반으로 개발모형의 전반적인 개발방향을 설정한 후, <그림 3-3>과 같은 개발과정을 통해 모형을 개발하였다.

첫째, 개발하고자 하는 모형의 가용자료(자료수집 환경, 자료의 양과 질)와 연산수행속도 등을 고려하여 기존연구 고찰을 수행하였으며, 대용량 자료조건에서 모형의 연산수행속도가 빠르면서 예측력이 우수한 ITS 예측분야 지식추출의 접근법으로 k -최근린 비모수회귀(k -nearest neighbor

non-parametric regression, 이하 KNN-NPR)을 선정하였다. 그리고 기존 KNN-NPR의 파라미터인 k 값(최인접 이웃의 개수)의 전역(Global) 최적화의 단점을 극복하기 위하여 모형의 수행과정에서 국부(Local) 파라미터 최적화과정을 통해 기존 전역 최적 k 의 문제점을 극복하고 예측모형의 정확도를 향상시켰다. 국부 k 값의 최적화는 군집화(Clustering) 모형의 개발을 통해 달성하였으며, 개발된 군집화 모형은 빠른 연산수행속도로 합리적인 군집을 안정적으로 분류/구축하였다.

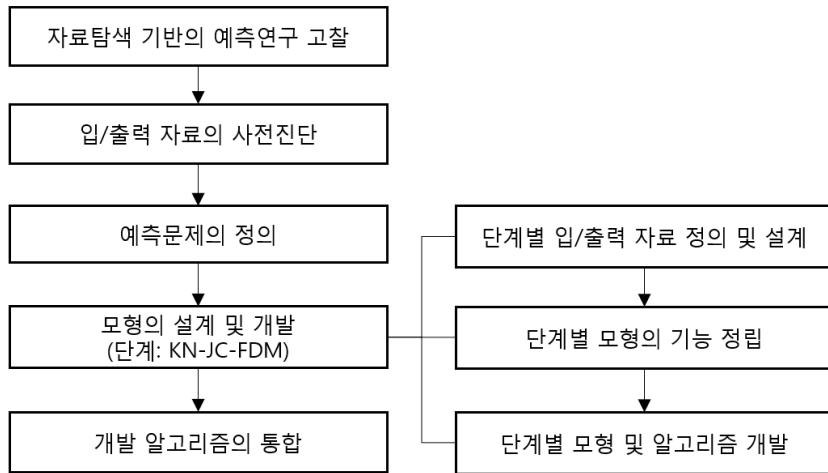
둘째, 예모형의 개발에 앞서 입·출력 자료의 사전진단(Prediagnosis)은 매우 중요하다. 개발모형의 최종 예측값인 거시적 소통상태의 시·공간적 행태는 입·출력 자료의 구조, 모형의 설계와 개발방식을 결정하기 때문이다. 본 연구에서 사전진단은 개발모형의 공간영역인 도로망의 차원에서 입력자료와 출력자료의 거시적인 시·공간 관계를 통해 수행되었다. 이는 개발모형의 경우 상태벡터를 이용하여 데이터에 내재된 동적인 미시적 행태 즉, 상태진화를 재구성(Reconstruction of state)하기 때문이다.

셋째, 예측문제의 정의는 예측모형의 최종 개발목표를 달성하고, 개발방향과 설계수준을 결정하는 핵심 요소이다. 이는 예측모형의 개발이 예측 문제를 풀기위한 일련의 과정이기 때문이다. 예측문제의 정의는 자료의 진단결과, 최종 결정된 입·출력 자료의 상태벡터, 중·장거리 경로통행시간 산정을 위해 필요한 예측영역을 고려하여 설정되었다.

넷째, 모형의 설계 및 개발은 기존연구의 고찰결과를 통해 1차 설정한 모형의 개발 접근법인 KNN을 구축하기 위한 KN단계, j 개의 의사결정 군집(Cluster)을 결정하기 위한 JC단계, 그리고 최종 의사결정 과정을 통해 장래 상태를 결정하는 예측 의사결정(Forecasting decision making, 이하 FDM)단계로 구성된다. 그리고 단계별 입·출력 자료의 정의와 설계를 통해 단계별 개발모형의 기능을 정립하였다. 단계별로 정립된 모형별 기능을 달성하기 위하여 단계별 모형과 알고리즘을 개발하였다.

마지막으로 개발 모형 및 알고리즘의 통합과정을 통해 대용량자료시스템에 탑재되어 장래 상태를 예측하는 KJC예측기의 원형(Prototype)을 개발하였으며, KJC 예측기는 단계별 목표를 달성하는 일련의 모듈형 모형이다.

따라서 모형의 구조변경이 용이하며 타 모형으로 모듈을 대체할 수 있기 때문의 시스템에 탑재된 후 모형의 구조변경이 용이하다는 장점이 있다.



<그림 3-3> 모형의 개발 과정

5) 개발모형의 구조

본 연구에서 개발된 KJC예측기는 경험적 지식을 기반으로 하는 상향식(Bottom-up) 의사결정모형으로서 KN단계, JC단계, 그리고 FDM단계로 구성되며, <그림 3-4>와 같다. KN단계는 이력 데이터베이스에 내재된 현행상태와 유사한 정보를 탐색/추출하는 지식탐색 역할을 수행함으로써 장래 교통상태를 결정하기 위한 k 개의 멤버(k 개의 유사상태 경험자)로 구성되는 1차 의사결정 집단을 구축한다. JC단계는 KN단계에서 구축된 의사결정 집단을 j 개의 2차 의사결정 집단(j 개의 군집)으로 세분화하는 역할을 수행하는 군집화(Clustering) 단계이다. FDM단계는 JC단계에서 결정된 j 개의 군집을 이용하여 장래 소통상태와 경로통행시간을 결정하는 최종 의사결정단계이다. 이상의 단계별 의사결정과정은 KJC예측기에서 개별 모듈로 구성되고, 개별 모듈은 입·출력자료로 연결된다.

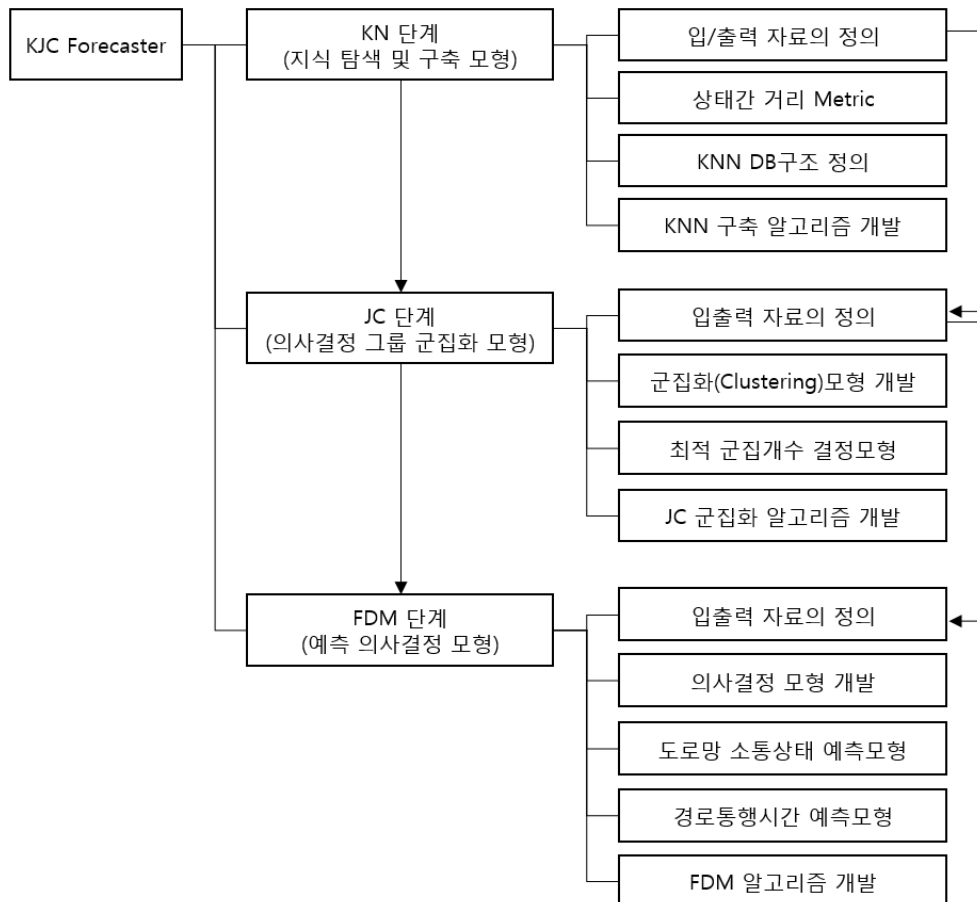
KN단계는 대용량 데이터베이스내에 존재하는 현행 교통상태와 유사한 과거 교통상태를 탐색·추출하여 장래상태를 설명하기 위한 1차 의사결정

집단인 k 개의 최근린 이웃(k -nearest neighbor, 이하 KNN)을 구성하며, 일종의 정보 추출 및 구축 모형이 이용된다. 이상의 과정을 달성하기 위하여 입·출력 상태벡터(State vector), 현행 상태와 과거 상태의 유사성을 설명하기 위한 상태거리 함수(State distance metric), 탐색된 유사성이 높은 과거 상태를 추출하고 KNN을 구축하기 위한 자료구조(Data structure)가 설정된다. 설정된 3개 구성요소는 KNN 추출 및 구축 모형을 구성하게 되며, 개발된 모형은 KNN구축 알고리즘으로 구현된다.

JC단계는 군집화(Clustering) 모형을 이용하여 KN단계에서 구축된 1차 의사결정 집단인 KNN을 j 개의 의사결정 집단으로 구성되는 2차 의사결정 집단들로 분류하며, 개발된 j -군집화 모형과 정적 군집개수 결정 모형이 이용된다. j -군집화 모형은 KNN을 j 개의 정적 군집으로 분류하는 패턴분류의 역할을 수행한다. j -군집화 모형은 절대 쌍(Absolute pair) 기반 접근법을 이용하여 1차로 구성된 하부군집(Subcluster)을 결합(Merging)하는 과정을 통해 군집의 수를 하부군집의 수에서 2개까지 줄여가는 결합형 군집화 모형이다. 그리고 최적의 군집수는 최적 군집수 결정모형에 의해 결정된다. 이상의 JC단계에서 j -군집화 모형과 최적 군집수 결정 모형은 JC 군집화 알고리즘으로 통합·구현된다.

FDM단계는 JC단계에서 구축된 j 개의 군집 즉, 2차 의사결정 집단들 중 최종 의사결정집단을 결정하고, 최종 의사결정집단을 구성하는 m 개 구성원($m < k$, $m > 1$)의 경험 즉, 과거 상태를 이용하여 도로망의 장래 소통상태를 결정한 후 경로통행시간을 예측하는 단계이다. 따라서 JC단계는 최종 의사결정집단 결정모형과 장래상태 예측모형으로 구성된다. 최종 의사결정집단 모형은 JC단계에서 결정된 군집들 중 현행상태와 유사성이 가장 높은 군집부터 낮은 군집까지 군집순위를 결정하고, 현행상태와 유사성이 가장 높은 군집을 최종 의사결정집단으로 결정한다. 장래상태 예측모형은 최종 의사결정집단의 m 개 구성원에 해당하는 과거상태와 KN단계에서 산정한 개별 구성원의 상태거리를 기반으로하는 유사성을 이용하여 장래상태를 산정한다. 장래상태의 예측은 도로망 소통상태 예측모형과 경로통행시간 예측모형으로 구성된다. 도로망 소통상태

예측모형은 m 개 의사결정 구성원의 과거 도로망 소통상태와 현행상태간의 유사성을 이용하여 도로망을 구성하는 모든 구간의 장래 소통상태를 추정한다. 경로통행시간 예측모형은 동적 구간기반 통행시간(Dynamic LPTT) 모형과 경로기반 경로통행시간(PPTT) 모형으로 구성된다. DLPTT모형은 구간별 장래 통행시간을 예측한 후 식(2-5)를 이용하여 경로통행시간을 산정하며, PPTT는 최종 의사결정집단의 구성원별 과거 경로통행시간과 유사성을 이용한 예측식으로 경로통행시간을 산정한다. 이상의 의사결정집단 선정모형, 도로망 소통상태 예측모형, 그리고 경로통행시간 예측모형은 FDM 알고리즘으로 통합·구현된다.

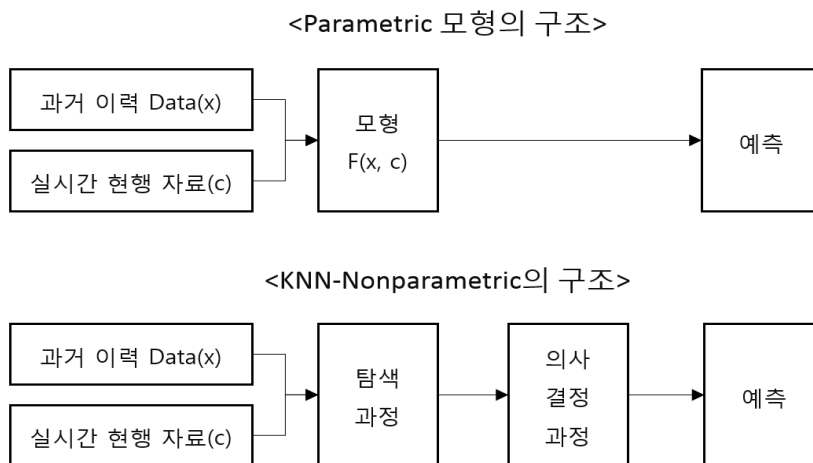


<그림 3-4> 개발모형(KJC Forecaster)의 구조

2. 데이터 탐색기반의 예측 연구

1) KNN-NPR 모형의 구조

파라미터 기반의 인위적인 모델링의 한계를 극복하기 위한 대안 중 하나가 비모수 회귀(Nonparametric regression, 이하 NPR)이며, NPR에서 k 는 k 개의 최근린 이웃(k -nearest neighbor, KNN)을 의미한다. NPR은 모수모형(Parametric model)과 다른 구조와 연산체계를 가지며, <그림 3-5>와 같다. 모수기반 모형은 일반적으로 확률 상태이론을 기반으로 개발된 인위적 모형을 이용하여 상태를 추정한다. 반면, NPR은 이력자료에 내재된 행태를 탐색과정을 통해 추출하고, 추출된 행태와 의사결정 과정을 이용하여 상태를 추정한다. 따라서 NPR은 (비록 혼돈 상태이론에 기초하지만) 특정 상태이론에 대한 가정이 필요하지 않으며, 다양한 특성을 가진 상태의 진화행태를 설명할 수 있다. NPR에서 KNN은 주어진 현행 상태벡터와 유사한 과거의 상태벡터 집단을 의미함과 동시에 예측문제에 있어 장래상태에 대한 의사결정 집단을 의미한다(Eubank, 1988). NPR은 모형 개발자에 의해 인위적으로 생성된 정보보다는 이력자료에 방대한 양의 경험적 지식이 내재되어 있다고 가정하고 이를 이용한다(Eubank, 1988).



<그림 3-5> 모수기반 모형과 NPR 모형의 구조

2) KNN-NPR의 이론적 배경과 적용

NPR은 혼돈 체계 이론(Chaotic system theory)에 기초하며, 이론적 원리는 혼돈 상태에서 패턴인식에 기인한다(Karlsson and Yakowitz, 1987; Davis and Nihan, 1991). Devijver(1982)는 시스템의 상태가 충분한 무작위 분포인 경우 즉, 입력자료의 개수(n) $\rightarrow\infty$ 이고 최인접 이웃의 개수(k)가 $k\rightarrow\infty$ 인 조건과 $k/n\rightarrow 0$ 을 만족하는 상태에서 NPR은 점근선적으로 최소 위험 결정(Asymptotically minimum risk decision)임을 밝혔다. 그리고 Yakowitz(1987)는 NPR을 시계열자료로 확대하였다. $n\rightarrow\infty$ 이고 $k\rightarrow\infty$ 조건의 $k/n\rightarrow 0$ 인 상태에서 KNN의 예측값은 최소 오차에 수렴하며, 수렴오차는 NPR 접근법 중 명백히 최적임을 밝혔다. 이러한 KNN 모형의 특성은 $k/n\rightarrow 0$ with $[n\rightarrow\infty, k\rightarrow\infty]$ 인 입력자료 조건에서 모수적 접근법과 유사하거나 우수한 추정력을 보일 수 있다. KNN은 입력자료 조건 때문에 방대한 양의 입력자료에 의해 독립변수와 종속변수를 설명하며, 사례기반 추론의 한 종류이다(Smith and Oswald, 2003).

예측분야에서 (모수)회귀모형은 독립변수 또는 입력상태로 설명되는 현재의 상태가 장래에도 동일하게 종속변수 또는 출력상태에 영향을 미친다는 가정에 대한 문제점이 있다(Oswald et al., 2000). 이에 비해 NPR기법은 현재 상태와 유사한 과거 상태를 탐색하여 장래 상태를 추정하는데 적용이 용이하다(Smith et al., 1999). NPR모형은 일반적인 회귀모형과 달리 입력과 출력자료간의 관계를 결정하는 모형의 구조를 가지게 됨으로 파라미터의 지속적인 재정산 없이 새로운 입력자료를 모형에 용이하게 추가할 수 있다. 그리고 자료의 통계적 구조에 대한 가정을 하지 않는 즉 Distribution-free의 장점이 있다.

NPR기법은 반복적 경험 즉, 과거의 패턴을 바탕으로 장래의 상태를 결정하는 의사결정 과정으로서 현행 상태와 유사한 과거의 상태로 구성되는 의사결정집단인 군집을 지속적으로 탐색하는 동적 군집모형이다. KNN-NPR의 군집을 정의하는 방법은 최인접 개수 k 를 정하는 KNN기

법과 최근접 개수의 \pm 범위를 설정하는 Kernel 이웃 기법이 있다.

KNN-NPR은 혼돈체계 이론에 기초하여 개발되었지만, 혼재된 상태 (Mixed situation)에 적용되었을 때 타 접근법에 비하여 명백한 장점을 가지며, 다양한 혼재된 동적상태의 예측에 적용되었다(Altman, 1992; Chang et al., 2010; Chang et al., 2011; Davis and Nihan, 1991; Devijver, 1982; Disbro and Frame, 1989; Guegan and Leroux, 2009; Karlsson and Yakowitz, 1987; Mulhern and Caprara, 1994). Robinson (1983), Mulhern and Caprara(1994)는 혼재된 무질서 상태에서 비선형 시계열 예측 문제를 NPR을 기반으로 모형화하였다. Disbro and Frame (1989), Mulhern and Caprara(1994)의 연구는 혼돈 상태에서 타 기법보다 NPR이 장점을 갖는다고 언급하였다. Karlsson와 Yakowitz(1987)은 NPR을 기반으로 강우량을 예측하였으며, Mulhern와 Caprara(1994)는 시장반응의 예측에 NPR을 적용하였다.

추가로 KNN-NPR 모형은 KNN 구축을 위한 탐색과정이 요구되기 때문에 연산수행 속도가 느리다는 단점이 있었다. 그러나 최근 컴퓨터의 연산기술 첨단화로 이러한 고질적 단점은 극복되었으며, 실시간 적용이 가능한 수준에 이미 도달하였다(Smith and Oswald, 2003, Chang et al., 2010).

3) ITS 예측분야의 KNN-NPR 적용 연구

NPR은 ITS 예측분야에서 거시적 교통변수의 예측과 추정에 주로 이용되었으며, 단일 시간대와 다중 시간대로 구분하여 고찰하면 다음과 같다. NPR은 단일시간대 교통량 예측연구에 주로 이용되었으며(Clark, 2003; Davis and Nihan, 1991; Kim et al., 2005; Lam et al., 2006; Smith and Demetsky, 1997; Smith et al., 2002; Smith and Oswald, 2003; Turochy, 2006), 이는 교통량의 시계열 변동이 타 교통변수에 비하여 심하기 때문이다. 그리고 Clark(2003)은 다중변수 예측에 KNN을 적용되었으며 통계기반 모형에 비하여 우수한 예측력을 보였다. 다중시

간대 연구로서 Smith and Demetsky(1996)은 고속도로의 교통량을 5분 시간대 길이(TL)로 4시간까지 예측하였다. Chang et al.(2012_b)은 고속도로 TG의 교통량을 15분 단위로 1시간까지 예측하였으며, 교통량이 급변하는 시점에서 TSA기법보다 우수한 정확도를 보였다. Yoon and Chang (2014)은 상태변화가 극심한 단속류 교통량의 예측에 NPR을 적용하였으며, 다중시간대 예측영역에도 불구하고 TSA기법의 단일 시간대 예측에 비해 매우 높은 예측 정확도를 보였다. 그리고 NPR은 결측자료 보정에 적용되었으며(장현호 등, 2010; Chang et al., 2012_a; Liu et al., 2008; Smith and Conklin, 2002; Smith et al., 2003), 확률적 상태이론을 기반으로 하는 ARIMA계열의 모형보다 낮은 추정오차를 보이는 것으로 보고되었다. 추가적으로 Qi and Smith(2004)는 방대한 양의 이력자료를 이용하여 유고지속시간을 예측하였다.

NPR은 경로통행시간 예측연구에도 적용되었으며, 통행시간의 길이가 긴 경우 타 모형에 비하여 예측력이 우수한 것으로 보고되었다(Chang et al., 2010; Chang et al., 2011; Bustillos and Chiu, 2011; Kim et al., 2011). Chang et al.(2010)은 노선버스의 주요 역간 출발지 기준 다중시간대 경로통행시간을 버스관리시스템의 경로통행시간 이력자료와 실시간 경로통행시간 자료를 이용하여 예측하였으며, 90분까지의 시간처짐을 극복하면서 장래 경로통행시간을 예측하였다. Chang et al.(2011)은 TCS 통행량의 KNN 패턴을 이용하여 중·장거리 고속도로 통행시간을 예측하였으며, 예측오차는 $\pm 5\%$ 이내로 우수하게 나타났다. Kim et al.(2011)은 루프 검지기 속도와 TCS 통행시간의 패턴매칭 기법을 이용하여 단/장거리 고속도로 통행시간을 예측하였다. Bustillos와 Chiu(2011)은 고속도로 유·출입 교통량과 N-Curve를 KNN과 결합하여 경로통행시간을 예측하였다. 이상의 연구들은 5~10% 이내의 예측오차를 보이면서 효과적으로 중·장거리 통행시간을 예측할 수 있음을 보였다. 반면 구간기반 경로통행시간(LPTT)이 아닌 경로기반 경로통행시간(PPTT)이라는 점에서 고정되지 않은 기·종점간 경로통행시간의 예측에는 한계가 있다.

3. 입력변수의 거시적 특성분석

1) 혼잡의 시·공간적 진화 형태

예측모형의 개발에 앞서 입·출력 변수의 형태 그리고 입력변수와 출력변수간의 관계를 사전진단(Predagnosis)하는 과정은 매우 중요하다. 이는 예측모형의 기반이 되는 상태이론의 선택, 모형의 성능과 목표의 설정, 그리고 입력변수와 모형의 예측성능을 사전에 파악할 수 있기 때문이다.

본 연구에서 개발되는 모형은 개별 구간기반이 아닌 거시적 도로망 차원에서 모든 구간의 소통상태와 통행시간을 예측한다. 따라서 네트워크 상에서 혼잡의 시·공간 진화와 패턴을 우선적으로 사전진단하였다. <그림 3-6>은 경부고속도로 대전→서울, 서울→대전 방향에 대한 혼잡의 시·공간적 진화를 5분 단위 구간평균속도로 24시간 동안 보여주고 있다. 혼잡은 지속적인 시·공간적 성장과정을 통해 거대 이동혼잡(Wide moving jam, 이하 WMJ), 일정 공간에서 장시간 지속되는 정적혼잡(Stationary jam, 이하 SJ), 교통사고로 인한 임의적 혼잡 등 다양한 형태로 나타났다. [참고: <그림 3-6>에서 혼잡의 기준은 구간통행속도(km/시)이며, 흑색에 가까울수록 모든 시간대(5분 간격 288개 시간대/일)에 대한 모든 구간의 속도 중 최저속도이고 백색에 가까울수록 최대속도이다.]

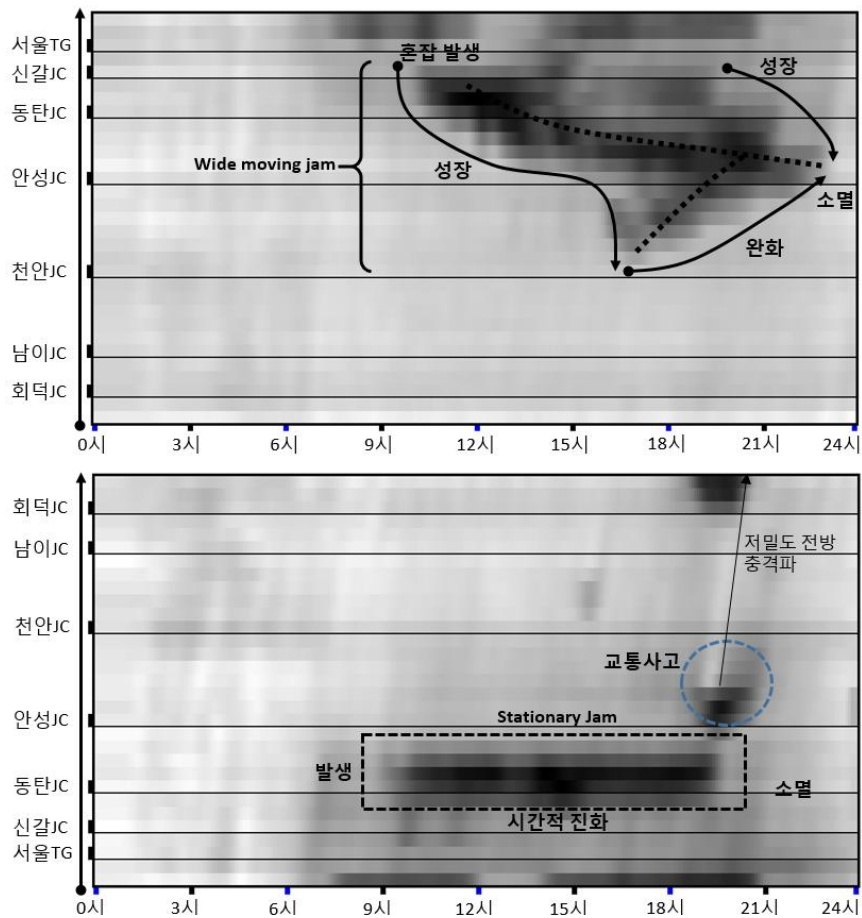
대전→서울 방향의 경우, 10시에 동탄→신갈JC의 구간에서 동시에 발생한 혼잡은 6시간에 걸쳐 천안JC까지 성장하면서 70km에 달하는 WMJ를 형성한 후, 16~20시까지 성장·둔화 동시에 보이고 있으며, 소멸과정을 거쳐 안성JC부근에서 23시에 소멸하고 있다. 이러한 WMJ의 발생·진화·소멸의 행태는 시·공간적으로 매우 다양하다. 그러나 시간적으로 해당 WMJ은 발생과 소멸까지 총 13시간 동안 진행되었으며, 공간적으로 충분한 시간을 가지며 성장한다는 점이다. 예측의 관점에서 볼 때, 이러한 혼잡의 시·공간적 전개과정 즉, 시간에 따른 소통상황은 도로망의 차원에서 충분히 설명이 가능할 것으로 판단된다. 이는 혼잡의 발생·성장·소멸의 진화과정에서 충분한 신호(Signal)가 교통변수의 시계열적 변동에 포함되어 있기 때문이다. 그리고 신갈-천안JC 구간에 혼잡이 집

중되는 이유는 지방부에서 출발한 통행이 서울방향으로 집중되기 때문이다. 이는 혼잡의 생애주기를 예측함에 있어 도로망 차원에서 통행량의 시간대별 방향성과 변화량을 신호로 이용할 경우 혼잡의 전개과정을 사전에 설명할 수 있음을 의미한다.

서울→대전 방향의 경우, 신갈JC→동탄JC 구간에서 9시에 동시에 발생한 혼잡은 20시까지 총 11시간동안 지속되고 있으며, 공간적으로 성장하지 않으면서 시간적 진화를 보이는 전형적인 SJ를 보여주고 있다. 이러한 혼잡의 특성은 동탄JC를 포함한 상류에서 출발한 통행이 지방방향으로 집중되기 때문이다. 일반적으로 SJ형태의 혼잡은 혼잡 지속시간 동안 지속적인 약한 성장과 쇠퇴의 과정을 반복하며, 이는 시계열 상태에서 변동으로 나타난다. 그리고 해당 혼잡은 상류부 혼잡상태의 변화에 따라 영향을 받고 있다. 서울TG 이전 혼잡의 상태 변화 그리고 양재IC→판교JC구간의 혼잡은 12~14시에 약화되고 있다. 상류부 혼잡상태의 변화는 해당 혼잡의 13~14시 동안의 혼잡 약화에 영향을 미치고 있으며, 이는 교통량 감소로 인한 국부적 현상으로 판단된다. 그리고 20시 이후 두 혼잡은 소거되고 있으며 양재IC→판교JC 구간의 혼잡이 오래 지속된 이유는 도시부 교통수요 특성이 강하기 때문으로 판단된다. 이상과 같이 도시부의 혼잡특성이 강한 구간에 대한 혼잡의 발생시점을 예측할 경우 오랜 시간에 걸쳐 지속되는 SJ의 특성상 보다 먼 장래 소통상태의 예측이 가능할 것으로 판단된다.

교통상태는 교통사고, 도로 점유공사 등 비 반복적 행태로 인한 영향을 받게 된다. 서울→대전 방향의 안성→천안JC 구간은 교통사고 발생에 의한 혼잡을 보여주고 있다. 19시에 발생한 사고의 여파는 20시 30분까지 지속되고 있으며, 사고처리 전까지 사고지점을 탈출한 교통량은 감소하고 있다. 따라서 비 반복적 상태로 인한 하류부의 자유교통류 상태가 시·공간 진화에 나타나고 있으며, 70km 하류부의 혼잡을 소거시키고 있다. 이와 같이 비 반복적 행태는 혼잡의 발생과 더불어 저밀도의 교통류를 하류부로 전파시키기 때문에 시·공간적으로 분리된 혼잡에 영향을

미친다. 이러한 비 반복적 행태는 구간기반 소통상태 예측에 있어 국부적이지만 해당 구간과 타 구간에 임의성을 발생시킨다. 따라서 본 연구에서는 구간기반 접근법에서 발생하는 임의성을 극복하기 위하여 도로망 기반의 전반적인 소통상태를 예측하도록 한다.



<그림 3-6> 혼잡의 시·공간 진화와 패턴

2) 프로브 통행량의 시·공간적 변화

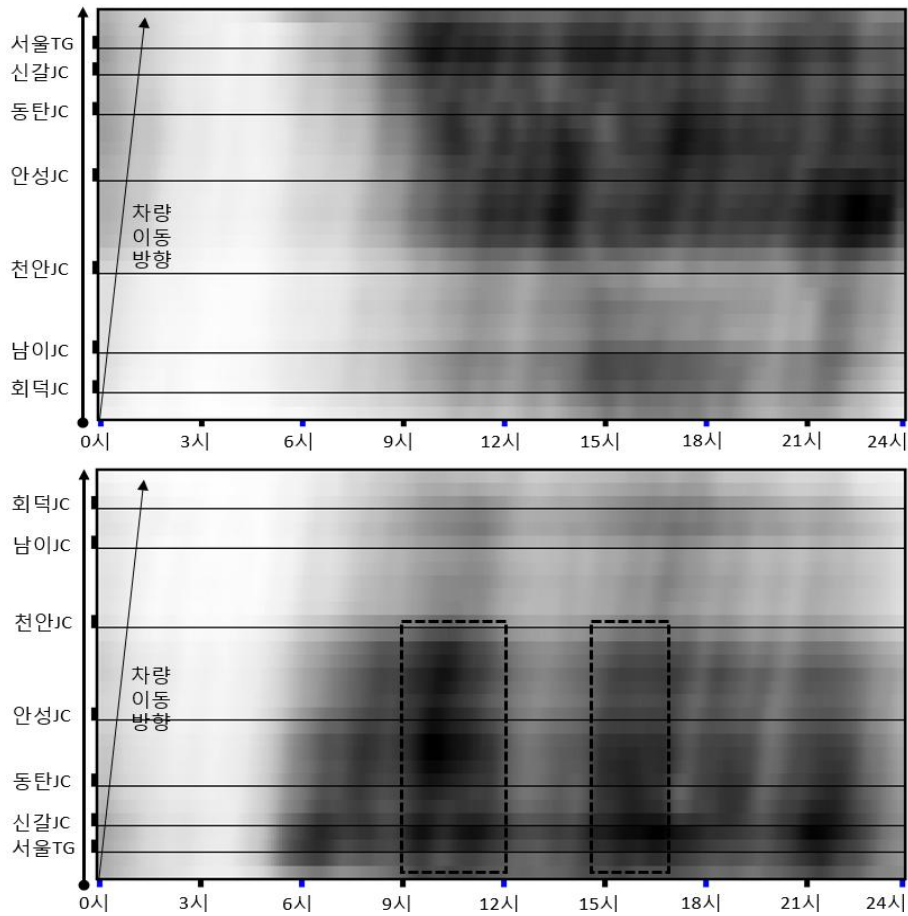
교통량의 시·공간 진화는 도로망에서 혼잡을 발생시킴과 동시에 소거시킨다. 그리고 구간 프로브 통행량은 교통량과 매우 높은 상관관계를

가지고 있다. 따라서 구간 프로브 통행량의 시·공간 진화행태를 입력변수로 이용하기 위한 사전진단을 수행하면 다음과 같다. <그림 3-7>은 경부고속도로 서울-대전 구간의 구간별 프로브 통행량에 대한 시·공간 상태를 방향별로 보여주고 있다. [참고: <그림 3-7>에서 구간 프로브 통행량(대/5분)은 흑색에 가까울수록 모든 시간대(5분 간격 288개 시간대/일)에 대한 모든 구간의 프로브 통행량 중 최대이고 백색에 가까울수록 최소이다].

대전→서울 방향의 프로브 통행량 변화는 9시 이후 급격히 증가한 후 시·공간적인 변동성을 보이기는 하지만 23시까지 일정하게 지속되고 있다. 프로브 통행량의 감소는 19시 회덕JC부터 감소하기 시작하여 24시 천안JC까지 감소하고 있지만, 천안JC이후부터는 24시 이후까지 지속되고 있다. 그리고 해당 노선의 경우 천안JC에서 유입 프로브 대수가 급격히 증가하는 특성을 보이고 있으며, <그림 3-6>의 서울방향의 WMJ와 밀접한 관련이 있는 것으로 판단된다. 따라서 대전→서울 방향의 프로브 통행량은 도로망의 하류부로 집중되는 특성을 보이며, 이러한 특성으로 <그림 3-6>에서와 같이 혼잡은 9시 이후에 집중되고 있다.

서울→대전 방향의 프로브 통행량의 경우, 단·중거리의 5시부터 시작되고 있으며 9시이후부터 중·장거리 통행이 증가하고 있다. 중·장거리 통행은 9~12시, 14~17시에 서울TG~천안JC구간에 집중되고 있다. 천안JC에서 통행량이 급격히 감소한 이유는 천안-논산 고속도로의 통행수요가 높기 때문으로 판단된다. 21시 전/후로 단·중거리 통행량이 증가한 후 23시에 급격히 감소하고 있다. 따라서 혼잡은 서울TG→천안JC구간에 오전 또는 오전과 오후에 집중되는 경향을 보일 것으로 판단된다.

이상과 같이 통행량의 시·공간적 진화행태 즉, 방향성과 변동량은 혼잡의 생애주기와 밀접한 관련을 가지는 것으로 판단된다. 따라서 통행량의 변화량과 방향성은 비 혼잡상태에서 혼잡의 발생 시점과 지점을 사전에 예측하고, 혼잡 발생이후 혼잡의 꼬리(Tail)에 도착하는 통행량의 변화와 혼잡의 머리(Head)를 탈출하는 통행량의 시·공간적 변화는 혼잡의 성장과 소멸을 예측하는데 있어 주요한 입력변수라고 판단된다.



<그림 3-7> 구간 프로브 통행량의 시·공간 변화

3) 프로브 통행량과 속도의 관계

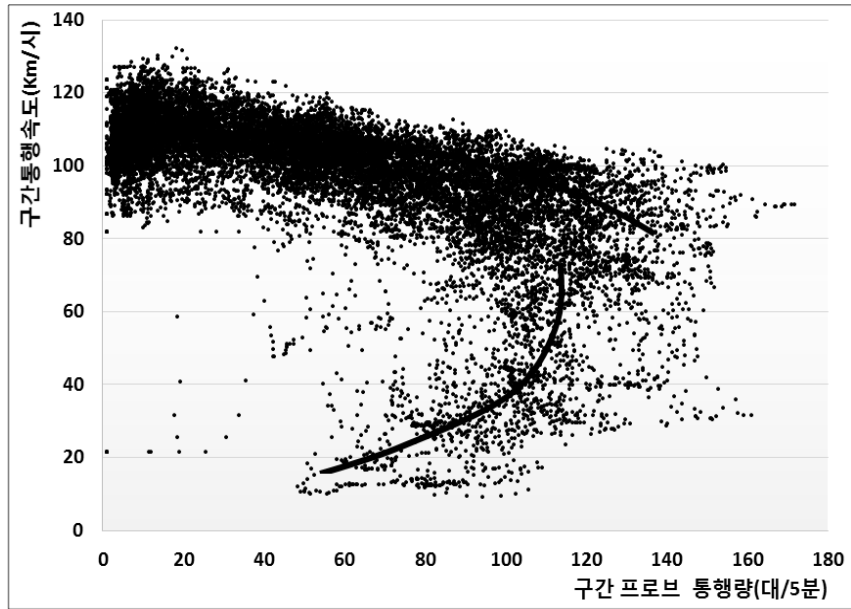
구간 프로브 통행량이 통행속도와 더불어 장래 상태에 대한 지표 (Indicator)의 역할을 수행하기 위해서는 통행량-속도의 관계를 보여야 한다. <그림 3-8>과 <그림 3-9>은 통행량-속도의 관계를 도로망 차원의 거시적 관계와 단일 구간에서 시간에 따른 미시적 관계로 보여주고 있다.

<그림 3-8>은 프로브 통행량이 속도를 설명하기 위한 최소한의 관계를 사전진단하기 위함이다. 프로브 통행량이 증가함에 따라 속도(km/시)는 115에서 70이전까지 감소하며, 용량상태는 대부분의 경우에 있어 120

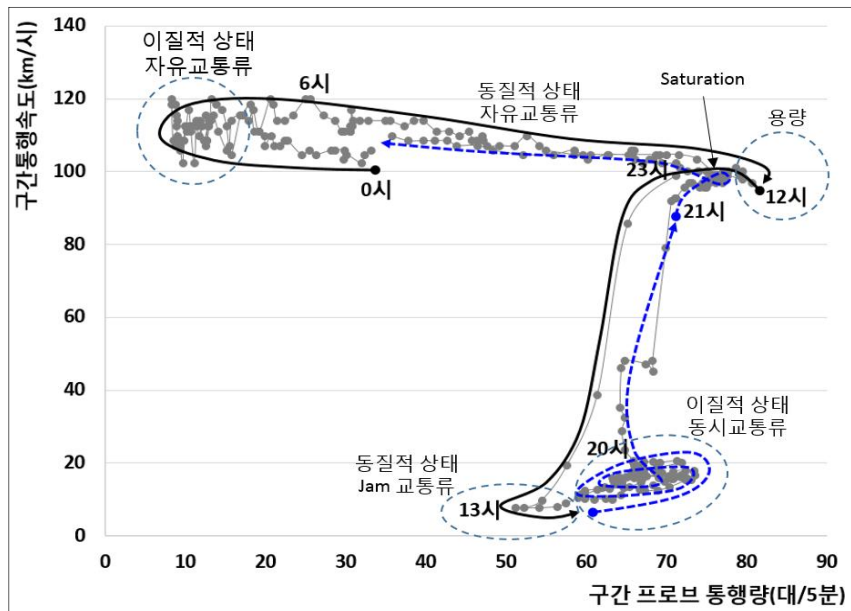
대/5분 이상을 유지하는 것으로 판단된다. 교통와해 이후 속도감소로 인한 구간 프로브 통행량은 50대/5분까지 감소하며 속도는 10이내로 급감하고 있다.

<그림 3-9>는 프로브 통행량-구간통행속도의 관계를 시간에 따라 보여주고 있다. [참고: 대상구간은 안성 휴게소 서울방향 구간이며, 구간 길이는 9.1km이다]. 구간통행속도이지만 지점 속도와 동일하게 정상교통류와 혼잡교통류간에 명확한 경계가 존재하고 있다. 정상교통류의 경우, 0~5시의 심야 및 새벽시간대에는 전형적인 자유교통류의 이질적(Non-homogeneous) 상태를 보인 후 6~12시까지 자유교통류 상태에서 동질적(Homogeneous) 상태가 나타나고 있다. 그리고 두 교통류간에는 (준)안정 상태가 존재한다. 용량상태는 12시 이후에 30분정도 지속되었으며, 동질적 상태의 혼잡교통류로 전환하는데 약 20분이 소요되었다. 동질적 혼잡교통류에서 가다서다(Stop-and-go) 현상과 동시(Syncronized) 교통류가 혼재하는 이질적 혼잡교통류로 30분후 전환되고 있으며 두 교통류간에 분명한 (준)안정상태가 존재하고 있다. 이질적 동시교통류는 13~20시까지 7시간동안 지속되고 있다. 20~21시 동안에 이질적 동시교통류는 포화교통류(Saturation flow) 상태에 도달하였으며, 1시간 동안 동시교통류 상태가 진행되는 동안 통행량 수준 70대/5분의 상태에서 속도는 18km/시에서 97km/시로 회복되었다. 포화교통류 상태는 2시간 동안 지속되었으며, 이후 동질적 자유교통류의 상태에서 다시 이질적 자유교통류 상태로 전환하고 있다. 따라서 프로브 통행량-속도의 관계는 역 λ 형태의 교통량-밀도 관계에서 전형적인 교통량-속도의 관계로 판단된다.

이상의 구간 프로브 통행량-구간통행속도 관계에서 교통상태가 경계 상태(Boundary condition)를 지나 다른 상태로 변화하는 현시전환(Phase transition)시 지점 교통류와 달리 20~30분의 시간이 소요되고 있다. 예측의 관점에서 이는 도로망 공간상에서 프로브 통행량-속도의 시간적 진화과정에 내재된 신호를 이용하면 교통류 상태의 현시전환 이전에 장래상태를 추정함에 있어 예측영역을 확장할 수 있음을 의미한다.



<그림 3-8> 구간 프로브 통행량과 구간통행속도의 거시적 관계

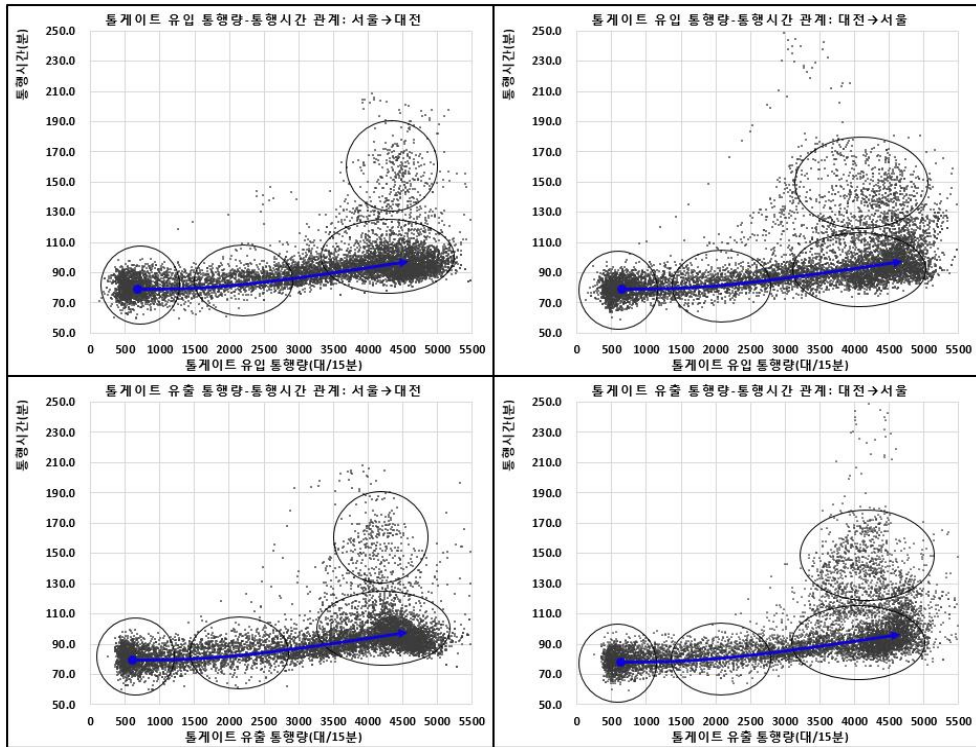


<그림 3-9> 구간 프로브 통행량과 구간통행속도의 미시적 관계

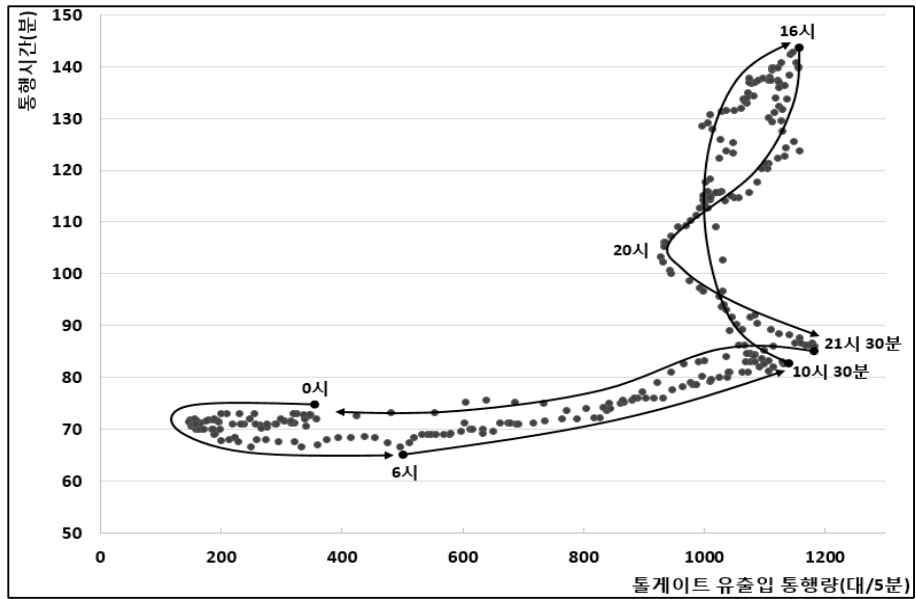
4) 유·출입 통행량과 경로통행시간의 관계

도로망의 소통상태는 도로망의 구간 소통상태와 현행 및 장래의 유·출입 통행량에 의해 변화한다. 유·출입 통행량은 연결점에서 방향성을 설명할 수 없지만, 전체 도로망의 측면에서 볼 때 유·출입 지점간의 경로 통행량에 대한 정보를 내재적으로 포함하고 있다. 예측의 관점에서 보면 도로망의 유·출입 통행량은 단시간에 도로망의 소통상태에 영향을 미치지 않는다. 그러나 구간 통행량과 속도로 설명력이 떨어지는 확장된 예측영역에 대하여 유·출입 지점간 출발 및 도착 통행의 집약된 신호로 이용될 수 있으며, 이를 이용한 고속도로 중·장거리 통행시간의 예측에 관한 연구가 보고되었다(Chang et al., 2011; Bustillos and Chiu, 2011).

<그림 3-10>은 경부고속도로 서울-대전구간의 서울TG와 대전IC간 방향별 경로통행시간과 TG 유·출입 통행량과의 거시적 관계를 각각 보여주고 있다. 유·출입 통행량은 70~90분의 통행시간 영역과 90분 이상의 통행시간 영역에 의해 양분되며, 통행시간은 유·출입 통행량 기준의 도로망 용량수준에서 2개의 영역으로 양분되고 있다. 이러한 관계를 시간적 변화에 따른 미시적 행태로 살펴보면 <표 3-11>과 같다. [참고: 경부선 서울-대전구간의 유·출입 통행량과 대전IC→서울TG 통행시간의 관계이다.] 10시 30분까지 증가한 통행량은 도로망의 용량수준에 도달후 혼잡이 발생하였다. 이로 인해 유출 통행량이 제약되면서 전체 통행량은 감소하지만 유입 통행량이 증가하면서 통행시간은 16시까지 5시간 30분에 걸쳐 증가하고 있다. 이후 20시까지 유·출입 교통량이 감소하면서 통행시간은 감소하며, 20시 이후 정체가 해소되어 감에 따라 유출 통행량의 증가로 통행량은 증가하고 있다. 특히 유·출입 교통량과 통행시간의 시간적 진화는 단 시간내에 이루어지지 않고 있으며, 이는 예측의 관점에서 충분한 시계열상의 신호로 이용할 수 있음을 의미한다. 따라서 유·출입 통행량이 구간 속도와 프로브 통행량의 시·공간적 상태와 동시에 고려될 경우, 혼잡의 생애주기를 설명할 수 있을 것으로 판단된다.



<그림 3-10> 유·출입 통행량과 경로통행시간의 거시적 관계



<그림 3-11> 유·출입 통행량과 경로통행시간의 미시적 관계

4. 개발모형의 예측문제 정의

1) 현행상태 입력변수의 선정

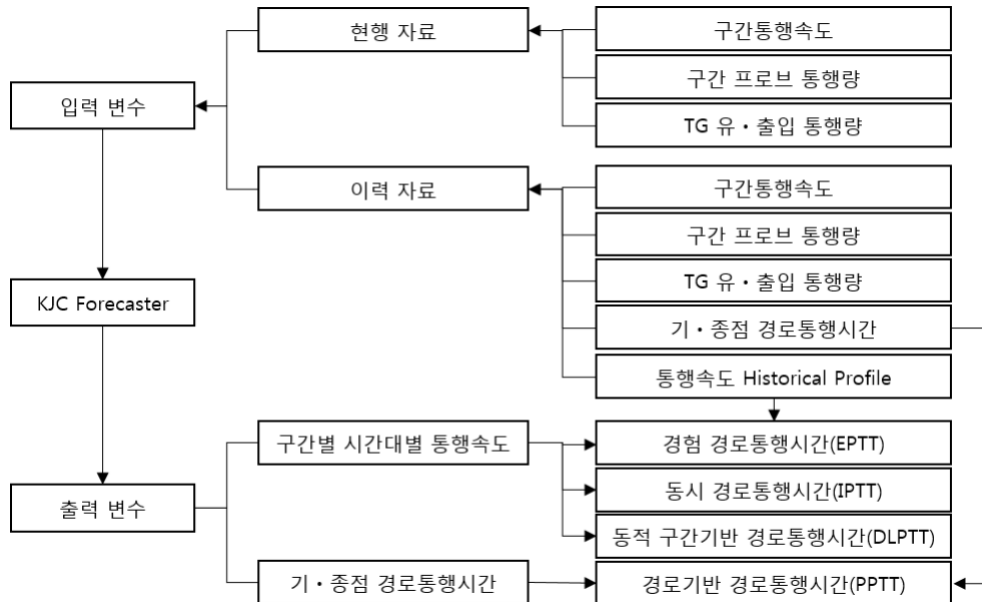
설명변수의 기술에 앞서, 장래 도로망의 소통상태를 100%의 정확도로 24시간 전에 구현하는 실시간 교통류 모의실험기가 있고, 모의실험결과를 데이터베이스에 지속적으로 저장하고 있다고 가정해 보자. 해당 교통류 모의실험기를 운영하기 위해서 필요한 과정은 매우 자명하다. 실시간 교통류 상태를 네트워크에 이식한 후, 실시간 및 장래의 통행량을 네트워크에 부하>Loading)해주어야 한다. 그렇다면 도로망 기반의 장래 예측을 위한 입력자료, 즉 설명변수는 무엇인지 또한 자명해진다. 예측모형이 실시간 소통상태와 통행량 상태의 시간적 진화과정에 내재된 신호(Signal)를 인지하고, 인지된 신호를 이용하여 입력자료에 포함된 과거의 모의실험 결과 즉, 입력자료를 이용하면 된다.

앞절에서 교통류 상태를 설명하기 위한 구간통행속도와 구간 프로브 통행량의 사전진단을 거시적 시·공간 진화의 행태로 수행하였으며, 진화의 과정에 장래 상태를 설명할 수 있는 신호가 내재되어 있음을 확인하였다. 그리고 통행속도와 프로브 통행량은 매우 밀접한 관련이 있음을 확인하였다. 유·출입 통행량의 경우 방향성은 없지만 방향성에 대한 통행량의 최종 결과물은 구간 프로브 통행량이다. 이와같이 통행량은 네트워크에 진입하면서 구간 통행량을 형성하고, 구간 통행량은 소통상태 즉, 통행속도에 영향을 미치며 혼잡을 형성하며, 발생한 혼잡은 유·출입 통행량에 영향을 미치게 된다. 이상의 관계에서 구간 프로브 통행량은 도로망 유·출입 교통량과 통행속도를 연결하는 매개변수(Intervening variable)라 할 수 있다. 따라서 예측모형의 현행상태 설명변수로 통행속도, 프로브 통행량 그리고 유·출입 통행량을 선정하였다. 추가적으로 본 연구에서 개발되는 KJC 예측기는 고속도로 뿐만 아니라 프로브가 수집되는 모든 도로의 소통상태를 예측대상으로 한다. 따라서 입력변수의 우선순위는 구간통행속도, 구간 프로브 통행량, 유·출입 통행량이다.

2) 입 · 출력 변수의 정의

예측모형의 장래 상태 추정을 위한 현행상태 입력변수가 설정되면 모형의 입 · 출력 변수를 설계해야 하며, 본 연구의 경우는 <그림 3-12>와 같다. 입력변수는 이력자료를 탐색하여 장래 상태를 추출하기 위한 현행 자료와 이력자료로 구성된다. 그리고 이력자료는 검색과 더불어 장래 상태의 후보군을 구성하기 위한 데이터베이스의 형태이다.

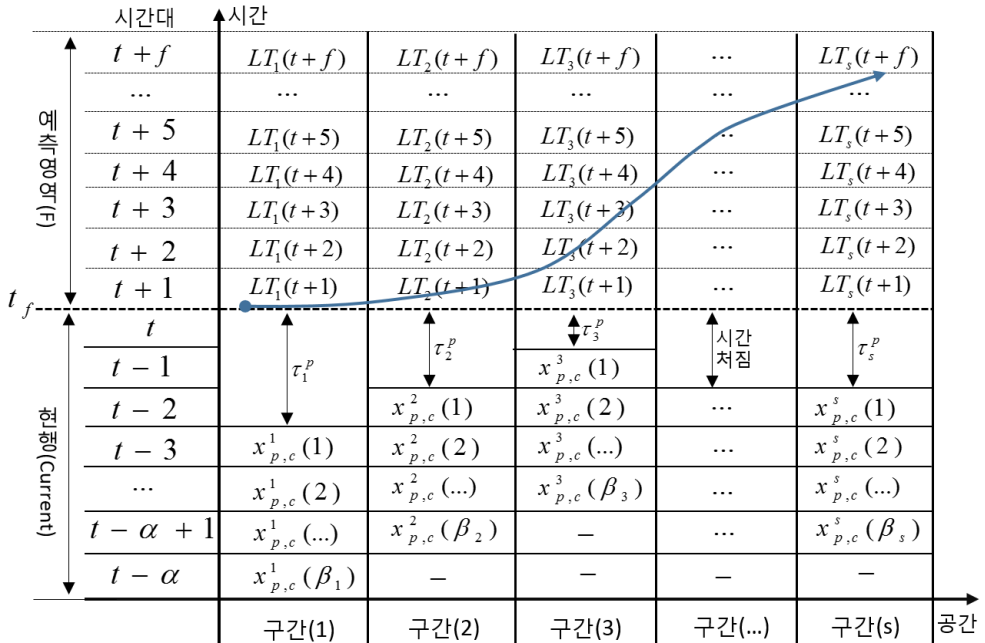
출력변수는 1차 출력변수와 2차 출력변수로 구성된다. 1차 출력변수는 도로망 소통상태인 구간별 시간대별 장래 통행속도와 이력 기 · 종점 경로통행시간이며, 2차 출력변수는 구간통행속도를 이용해 산정되는 3개 구간기반 경로통행시간[EPTT, IPTT, DLPTT]과 이력 기 · 종점 경로통행시간으로 산정되는 경로기반 경로통행시간[PPTT]이다. EPTT는 시점 $t+1$ 에서 구간통행시간과 시점 $t+1$ 이후의 이력속도 Profile을 이용하여 산정되며, IPTT는 시점 $t+1$ 에서 구간통행시간으로 산정된다. DLPTT는 예측모형의 예측영역내의 구간통행시간으로 산정되고, PPTT는 추출된 이력 경로통행시간자료를 이용하여 산정된다.



<그림 3-12> 개발모형의 입 · 출력 자료구조

3) 예측 문제의 정의

예측의 시·공간을 정의한 후 예측 문제를 정의하도록 하며, 예측의 시·공간은 <그림 3-13>과 같다. 예측의 공간적 범위는 도로망을 구성하는 속도자료가 수집되는 모든 구간($S=[1,2,...,s]$)을 대상으로 한다. 그리고 현행 자료의 변수집합(p)은 구간 및 지점 수집체계 정보를 모두 포함한다. 구간수집체계를 통해 수집되는 현행 자료는 자료수집에 시간처점이 발생하며, 시간처점의 길이도 다르기 때문에 실시간 적용을 위해서는 시간처점을 고려해야 한다. 따라서 구간(s)별 수집자료(p)별 시간처점(τ_s^p)을 고려한 도로망(S) 차원의 현행(Current) 상태벡터를 $X_{p,c}^s$ 로 정의한다. 시간대 길이(TL)로 구분된 장래 시간대의 개수인 예측영역($F=[1,2,...,f]$)은 식(2-4)의 조건을 충족하도록 구간통행시간($LT_s(t+f)$ $\forall s, f$)을 예측하고, 경로통행시간을 산정하도록 확장된다.



<그림 3-13> 개발모형의 시·공간 정의

본 연구의 예측문제는 개발모형의 Input-Output의 관계인 다중 변수 - 다중 변수의 입·출력 구조를 기반으로 식(3-1)과 같이 정의되며, 현행 상태의 정보를 이용하여 장래 상태의 구간 통행속도와 통행시간을 예측하고, 예측된 구간통행시간을 이용하여 기·종점간 경로통행시간을 예측함에 있어 불확실성 즉, 예측오차를 최소화하는 문제이다.

다중 입력체계인 현행 상태는 도로망을 구성하는 구간집합(S)과 구간의 다중 입력변수(p)에 대한 시간대(t)에서 현행 상태벡터 $X_p^s(t)$ 로 구성된다. 그리고 개발모형의 실시간 적용을 위하여 시간대(t)에서 Embedding size인 상태의 시계열자료 개수(l)는 p 별 구간(s)별 자료의 시간처짐(τ_p^s)을 고려하게 된다. 따라서 현행 상태 입력체계는 입력변수의 개수에 대한 제한이 없으며, 용이하게 입력변수의 구조를 조정할 수 있다.

예측은 하위 예측대상인 도로망 소통상태(구간통행속도, 구간통행시간)과 이를 이용한 상위 예측대상인 경로통행시간 예측으로 구성된다. 하위 예측대상인 구간 통행속도와 통행시간($LS_s(t+f), LT_s(t+f), \forall s, f, f \in F$)는 도로망을 구성하는 모든 구간(s)의 모든 예측영역 $F=[1, 2, \dots, f]$ 에 대하여 수행된다. 그리고 상위 예측대상인 경로통행시간은 도로망의 기·종점(o, d)에 대한 4개의 경로통행시간으로 구성된다. 4개의 통행시간은 IPTT(동시 경로통행시간)= $f'(LT_s(t+1))$, EPTT(경험 경로통행시간)= $f'(LT_s(t+1), LT_s^{\text{expr}}(t+f))$ $f > 1$, DLPTT(동적 구간기반 경로통행시간)= $f'(LT_s(t+f))$, PPTT(경로기반 경로통행시간)= $f'(PT_{o,d}^{\text{expr}}(t+1))$ 로 정의된다. 여기서, $LT_s^{\text{expr}}(t+f)$ 는 이력속도 Profile로 설명되는 $t+f$ 시점에서 과거 구간통행시간이며, $PT_{o,d}^{\text{expr}}(t+1)$ 는 이력자료로 설명되는 $t+1$ 시점에서 o, d 간 과거 경로통행시간이다.

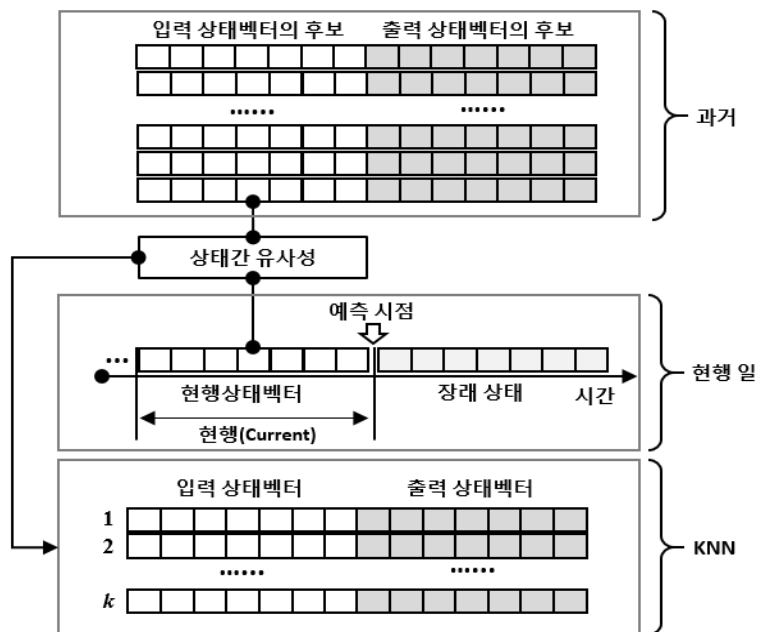
Given $X_p^s(t)$ with l, τ_p^s 식(3-1)

Predict $\begin{cases} LS_s(t+f), LT_s(t+f), \forall s, f \\ PT_{o,d}(t+1), \forall o, d \end{cases}$ with min. of uncertainty

5. KN단계 모형의 개발

1) 상태벡터의 정의

KNN 구축을 위한 상태벡터의 정의에 앞서 상태벡터와 KNN의 개념을 일차원 상태벡터로 설명하면 <그림 3-14>와 같다. 상태벡터는 현행 상태벡터, 입력 상태벡터, 그리고 출력 상태벡터로 구성된다. 현행 상태벡터는 예측 시점 이전의 시계열 자료를 의미한다. 그리고 KNN은 k 개의 입력 상태벡터와 출력 상태벡터로 구성된다. KNN의 입력 상태벡터는 입력자료에 포함된 입력상태 후보 중 상태간 유사성을 이용하여 선정된 현행 상태벡터와 유사한 k 개의 선택된 후보이며, 출력 상태벡터는 선택된 입력 상태벡터의 과거 출력 상태벡터이다.



<그림 3-14> 상태벡터와 KNN의 개념

KN단계는 KNN 접근법을 기반으로 현행 상태벡터와 데이터베이스에 존재하는 유사한 과거의 k 개 상태벡터를 탐색하여 k 개의 이웃을 구축하는 자료추출의 역할을 수행하는 단계로서 KNN을 구축하기 위해서는 상태벡

터의 정의가 선행되어야 한다. 상태벡터는 현행(Current) 상태를 설명하기 위한 현행 상태벡터(X_c), 현행 상태의 탐색 대상이 되는 과거 상태벡터(X_h) 그리고 X_h 의 미래에 해당하는 과거상태인 출력 상태벡터(O)로 정의한다. 출력상태 벡터(O)는 통행속도 출력 상태벡터(O_v)와 경로통행시간 출력 상태벡터(O_{pt})로 구성되며, $O=[O_v, O_{pt}]$ 로 정의한다.

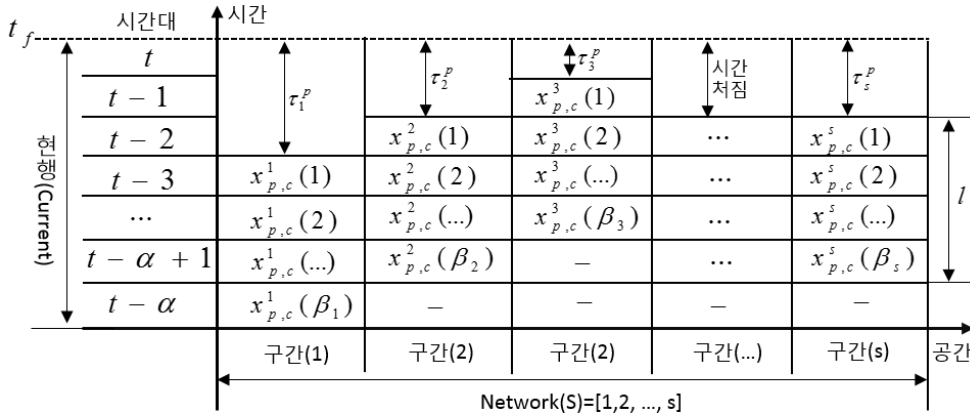
상태벡터의 설명에 앞서, 필요한 변수를 정의하면 <그림 3-15>와 같다. 도로망은 s 개로 구성되는 일련의 구간집합($S=[1,2, ...,s]$)으로 정의한다. 현행 상태벡터를 구성하는 입력변수는 프로브의 구간통행속도(kph), 구간 프로브 통행량(대/ TL), 그리고 도로망 유·출입 통행량(대/ TL)인 3개로 구성되며, $p=[\text{구간통행속도}, \text{구간 프로브 통행량}, \text{유·출입 통행량}]=[v, b, q]$ 로 정의한다.

개발모형을 실시간으로 적용하기 위해서는 입력자료의 수집에 발생하는 시간처짐을 고려해야 한다. 시간처짐은 p 별 s 별로 예측시점(t_f)을 기준으로 고려되며, τ_p^s 로 정의한다. 구간속도와 프로브 통행량은 동일한 시간처짐이 발생함으로 시점 t_f 에서 구간(s)의 변수(p)에 대한 시간처짐(τ_s^p)은 시점 t 에서 링크(s)의 통행시간($LT_s(t)$, 분)과 시간대의 길이(TL)를 이용하여 식(3-2)와 같이 정의한다. 이는 구간검지체계의 경우, 구간의 통행시간 만큼 시간처짐이 발생하기 때문이며, 시간처짐의 값을 1추가 함으로써 Outlier를 제외한 모든 구간통행이 완료된 경우를 안정적으로 고려하게 된다. 그리고 $LT_s(t)$ 와 TL 의 시간적 차원은 동일하다고 정의한다. 추가적으로 TG 유·출입 통행량과 같은 지점검지 방식의 경우는 $LT_s(t)=0$ 임으로 시간처짐은 1로 고정되며, 해당 자료를 수집하는 시스템별 집계시간 주기로 설정할 수 있다. 본 연구의 경우, 지점검지 방식의 시간처짐은 1로 가정한다.

$$\tau_s^p = LT_s(t)/TL+1, \forall p, s \quad \text{식(3-2)}$$

이상의 τ_s^p 을 이용하여 예측시점(t_f)에서 현행 시계열 상태의 개수, 즉 Embedding size(l)를 고려하면 p 별 s 별 시계열 상태의 시간공간은 시간

대 $t - \tau_s^p$ 에서 $t - \tau_s^p - l + 1$ 이다. 본 연구에서 현행 및 이력 상태벡터는 p 별 s 별 시계열 상태를 고려한 도로망(S) 기반의 p 별 상태로 정의한다. 따라서 상태벡터를 효율적으로 설명하기 위하여 τ_s^p 를 고려한 l 은 $[1, 2, \dots, \beta_s^p]$ 로 정의한다. 따라서 l 로 구성되는 일련의 시간대 $[t - \tau_s^p, t - \tau_s^p - 1, t - \tau_s^p - 2, \dots, t - \tau_s^p - l + 1]$ 는 $[1, 2, 3, \dots, \beta_s^p]$ 로 정의된다.



<그림 3-15> 시간처짐을 고려한 상태벡터의 시·공간 정의

본 연구의 예측모형은 3개 입력변수($p = [\text{구간통행속도}, \text{구간 프로브 통행량}, \text{유·출입 통행량}] = [v, b, q]$)를 이용한다. 따라서 3개 입력변수에 대한 상태벡터(X_p)가 정의되어야 하며, 다음과 같다.

구간통행속도(v)와 구간 프로브 통행량(b)의 현행 상태벡터 $X_{p,c}$, $p = \{v, b\}$ 의 차원은 $l \times S$ 로 동일하게 정의하며, 입력변수($p = \{v, b\}$)에 대한 상태벡터의 개별 상태 즉, 구간(s)별 시간대(β_s^p)별 상태는 $x_{p,c}^s(\beta_s^p)$ 로 정의한다. 이를 풀어쓰면 식(3-3)과 같다.

$$X_{p,c} = \begin{bmatrix} x_{p,c}^1(1), & x_{p,c}^2(1), & \dots, & x_{p,c}^s(1) \\ x_{p,c}^1(2), & x_{p,c}^2(2), & \dots, & x_{p,c}^s(2) \\ \dots, & \dots, & \dots, & \dots \\ x_{p,c}^1(\beta_1^p), & x_{p,c}^2(\beta_2^p), & \dots, & x_{p,c}^s(\beta_s^p) \end{bmatrix} \quad \text{식(3-3)}$$

유·출입 통행량 현행 상태벡터($X_{q,c}$)의 시간적 차원은 l 로 동일하다. 그러나 공간적 차원은 유·출입 지점의 개수에 양방향(In/Out) 통행량이 존재함으로 $g=2 \times \text{TG}$ 개수로 정의한다. 따라서 $X_{q,c}$ 의 차원은 $l \times g$ 이다. 상태벡터의 개별 상태 즉, 지점(g)별 시간대(β_g^t)별 상태는 $x_{q,c}^g(\beta_g^t)$ 로 정의되며, 이를 풀어쓰면 식(3-4)와 같다.

$$X_{q,c} = \begin{bmatrix} x_{q,c}^1(1), x_{q,c}^2(1), \dots, x_{q,c}^g(1) \\ x_{q,c}^1(2), x_{q,c}^2(2), \dots, x_{q,c}^g(2) \\ \dots, \dots, \dots, \dots \\ x_{q,c}^1(\beta_1^t), x_{q,c}^2(\beta_2^t), \dots, x_{q,c}^g(\beta_g^t) \end{bmatrix} \quad \text{식(3-4)}$$

과거 상태벡터의 정의에 앞서 이력상태의 집합을 H 로 정의하며, H 를 구성하는 개별 과거상태를 h , $h \in H$ 로 정의한다. 따라서 과거상태는 H 개로 구성되며, 모든 h 는 현행상태의 시간대에 종속적이라고 정의한다. 이러한 시간적 종속성(Dependency)은 방대한 양의 데이터베이스에 대한 탐색시간을 1/시퀀스(개수/일)로 줄여줌으로써 모형의 연산수행 속도를 증가시킨다. 예를 들어 288시퀀스/일의 데이터베이스 구조를 가지는 경우, 데이터 검색량은 1/288로 감소한다(Chang et al., 2012b).

h -번째 과거 상태벡터($X_{p,h}$, $p = \{v, b\}$)는 $X_{p,c}$ 와 동일 차원이며, 시간적으로 $X_{p,c}$ 에 종속적이다. 따라서 구간통행속도와 구간 프로브 통행량의 과거 상태벡터를 풀어쓰면 식(3-5)와 같다.

$$X_{p,h} = \begin{bmatrix} x_{p,h}^1(1), x_{p,h}^2(1), \dots, x_{p,h}^s(1) \\ x_{p,h}^1(2), x_{p,h}^2(2), \dots, x_{p,h}^s(2) \\ \dots, \dots, \dots, \dots \\ x_{p,h}^1(\beta_1^p), x_{p,h}^2(\beta_2^p), \dots, x_{p,h}^s(\beta_s^p) \end{bmatrix} \quad \text{식(3-5)}$$

h -번째 유·출입 교통량 과거 상태벡터($X_{q,h}$)의 시·공간적 차원은 $X_{q,c}$ 에 종속적이며, 식(3-6)과 같다.

$$X_{q,h} = \begin{bmatrix} x_{q,h}^1(1), x_{q,h}^2(1), \dots, x_{q,h}^g(1) \\ x_{q,h}^1(2), x_{q,h}^2(2), \dots, x_{q,h}^g(2) \\ \dots, \dots, \dots, \dots \\ x_{q,h}^1(\beta_1^q), x_{q,h}^2(\beta_2^q), \dots, x_{q,h}^g(\beta_g^q) \end{bmatrix} \quad \text{식(3-6)}$$

현행 상태벡터와 과거 상태벡터가 정의되면, 현행 상태벡터와 과거 상태벡터간의 유사성(Similarity)을 이용하여 추출되는 정보인 출력 상태벡터를 정의해야 한다. 출력상태 벡터(O)는 구간통행속도(v) 출력 상태벡터(O_v)와 경로통행시간(pt) 출력 상태벡터(O_{pt})로 구성된다. O_v 는 시간대($t+1$)에서 시간대($t+f$)까지의 모든 구간(S)에 대하여 통행속도를 출력하며, 시·공간 차원은 $f \times s$ 로 정의한다. 따라서 h -번째 과거 상태벡터에 대한 $O_{v,h}$ 의 s 별 $t+f$ 별 상태는 $x_{v,h}^s(t+f)$ 로 정의할 수 있으며, 이를 이용하여 $O_{v,h}$ 를 풀어쓰면 식(3-7)과 같다.

$$O_{v,h} = \begin{bmatrix} x_{v,h}^1(t+1), x_{v,h}^2(t+1), \dots, x_{v,h}^s(t+1) \\ x_{v,h}^1(t+2), x_{v,h}^2(t+2), \dots, x_{v,h}^s(t+2) \\ \dots, \dots, \dots, \dots \\ x_{v,h}^1(t+f), x_{v,h}^2(t+f), \dots, x_{v,h}^s(t+f) \end{bmatrix} \quad \text{식(3-7)}$$

마지막으로 O_{pt} 는 대상 도로망(S)에 존재하는 존(Zone)으로 설정된 IC 또는 TG간의 경로통행시간 매트릭스이다. 존 집합(Z)은 z 개의 존으로 구성되며, $Z = \{1, 2, \dots, z\}$ 로 정의한다. O_{pt} 는 존간의 경로통행시간으로 구성되며, 차원은 $z \times z$ 이다. 따라서 h -번째 과거 상태벡터에 대한 시점 $t+1$ 에서 O_{pt} 는 출발지 존($o, o \in Z$)에서 도착지 존($d, d \in Z$)까지 개별 경로통행시간 $o_h^{o,d}$ 로 구성되며, 식(3-8)과 같다.

$$O_{pt,h} = \begin{bmatrix} o_{pt,h}^{1,1}(t+1), o_{pt,h}^{1,2}(t+1), \dots, o_{pt,h}^{1,z}(t+1) \\ o_{pt,h}^{2,1}(t+1), o_{pt,h}^{2,2}(t+1), \dots, o_{pt,h}^{2,z}(t+1) \\ \dots, \dots, \dots, \dots \\ o_{pt,h}^{z,1}(t+1), o_{pt,h}^{z,2}(t+1), \dots, o_{pt,h}^{z,z}(t+1) \end{bmatrix} \quad \text{식(3-8)}$$

2) 상태간 거리 Metric

현행 상태벡터($X_{p,c}$)와 유사한 과거 상태벡터($X_{p,h}$)를 탐색하여 k 개의 출력 상태벡터(O)로 구성된 군집의 구축은 탐색과정을 기반으로 하는 KNN구축 알고리즘을 통해 수행되며, 탐색의 과정에서 상태간 유사성(Similarity)을 이용한다. 일반적으로 상태간 유사성을 설명하기 위하여 상태 거리, 통계적 검증 등 다양한 기법이 이용된다. 본 연구에서는 Minkowski distance metric로 알려진 L_m 거리(L_m metric)를 이용하여 상태간 유사성을 산정하도록 한다. L_m 거리는 $m=1$ 인 경우 Manhattan Distance, $m=2$ 이면 Euclidean Distance로 알려져 있으며, 식(3-9)와 같다.

$$L_m = L_m(x, y) = \left(\sum_{i=1}^d |x_i - y_i|^m \right)^{1/m}, \quad d \geq i \quad \text{식(3-9)}$$

교통류 상태 시계열자료는 상태 예측에 주요한 신호인 소음(Noise)을 포함하고 있다. 이러한 소음은 장래 상태의 예측에 있어 주요한 신호이며, L_m 거리는 이러한 소음 즉, 신호에 민감하다(Chang et al., 2012b, Yoon and Chang, 2014). 따라서 본 연구에서는 상태벡터간 차이가 증가할수록 상태거리가 적정 수준에서 증가하는 Euclidean거리(이하 UD)를 이용하도록 한다.

본 연구에서 UD 는 ①현행 상태벡터와 유사한 이력 상태벡터를 탐색하는 과정과 ②KNN 군집구축 알고리즘의 수행과정에서 입력변수 집합(p)에 대한 전체거리를 이용하여 가장 거리가 긴 이웃(Longest) 즉, 과거의 상태를 KNN 군집에서 제거하기위해 이용된다.

이력자료 탐색과정에서 UD 는 입력변수(p)에 대한 현행 상태벡터($X_{p,c}$)와 유사한 h 번째 과거 상태벡터($X_{p,h}$, $h \in H$)간의 상태거리(u_h^p)를 산정하며, 식(3-10)과 같이 정의할 수 있다. 이를 다시 풀어쓰면 식(3-11)과 같다.

$$u_h^p = L_2(X_{p,c}, X_{p,h}) \quad \text{식(3-10)}$$

$$u_h^p = \left[\sum_{i=1}^{\beta_n^p} \sum_{j=1}^{n=\{s,g\}} \left| x_{p,c}^j(i) - x_{p,h}^j(i) \right|^2 \right]^{1/2} \quad \text{식(3-11)}$$

일반적으로 ITS 시스템 자료는 다양한 형태의 누락자료를 포함하고 있다. 이러한 누락자료는 상태공간을 왜곡함으로 상태거리에 의한 유사성을 담보할 수 없다. 누락정보를 고려한 상태거리의 산정과 조정방법에 앞서, 본 연구에서는 모형의 구조화와 연산과정을 효율적으로 설명하기 위하여 행렬의 곱(\cdot)에 대한 연산을 식(3-12)와 같이 정의하며, 이는 실제 알고리즘의 구현에 있어 매우 효과적이다.

$$A \cdot B = \begin{pmatrix} a, b \\ c, d \end{pmatrix} \cdot \begin{pmatrix} 1, 2 \\ 3, 4 \end{pmatrix} = \begin{pmatrix} a \times 1, b \times 2 \\ c \times 3, d \times 4 \end{pmatrix} \quad \text{식(3-12)}$$

누락자료는 현행 상태벡터와 이력 상태벡터에서 모두 발생한다. 따라서 p 에 대한 현행(Current) 상태의 누락정보 상태(M_c^p)와 h 번째 과거 상태벡터의 누락정보 상태(M_h^p)를 정의하며, 차원은 해당 p 의 상태벡터와 동일하다. 누락정보 상태는 Binary이며, [정상=1, 누락=0]으로 정의한다. 예를 들어 $X_{v,c}$ 에 누락자료가 포함된 경우, M_c^v 는 식(3-13)과 같다.

$$IF \begin{bmatrix} x_{v,c}^1(1), & -, & ..., & x_{v,c}^s(1) \\ x_{v,c}^1(2), & x_{v,c}^2(2), & ..., & - \\ ..., & ..., & ..., & ... \\ x_{v,c}^1(\beta_1^v), & -, & ..., & x_{v,c}^s(\beta_s^v) \end{bmatrix} THEN M_c^v = \begin{bmatrix} 1, & 0, & ..., & 1 \\ 1, & 1, & ..., & 0 \\ ..., & ..., & ..., & ... \\ 1, & 0, & ..., & 1 \end{bmatrix} \quad \text{식(3-13)}$$

따라서 p 에 대한 전체 누락정보 상태($M^p, \forall p$)는 현행 상태벡터와 과거 상태벡터의 누락정보 상태인 M_c^p 와 M_h^p 그리고 식(3-14)를 이용하여 식(3-14)와 같이 산정된다. 그리고 입력변수 p 에 대해 양호한 자료의 구성비($r_p, \forall p, 0.0 \sim 1.0$)는 식(3-15)를 이용하여 산정된다.

$$M^p = M_c^p \cdot M_h^p \quad \text{식(3-14)}$$

$$r_p = \left[\sum_{i=1}^{n=\{s,g\}} \sum_{j=1}^{\beta_n^p} M^p(i,j) \right] \times (l \times n)^{-1} \quad \text{식(3-15)}$$

이상에서 산정한 M^p 와 r_p 를 이용하여 비정상 상태의 조정된 현행 상태벡터($X_{p,c}$)와 유사한 h 번째 과거 상태벡터($X_{p,h}$, $h \in H$)간의 상태거리(u_h^p)는 식(3-16)과 같이 계산된다.

$$u_h^p = \left[\sum_{i=1}^{\beta_n^p} \sum_{j=1}^{n=\{s,g\}} r_n^{-1} \times \left(M_p(i,j) \cdot |X_{p,c}^j(i) - X_{p,h}^j(i)|^2 \right) \right]^{1/2} \quad \text{식(3-16)}$$

KNN 군집구축 알고리즘의 수행과정에서 초기 k 개의 군집이 구축되면, 이후에 탐색되는 h 번째 과거 상태벡터의 출력 상태벡터는 기 구축된 KNN중 현행 상태벡터와 거리가 가장 긴 이웃(Longest)의 상태거리보다 작은 경우 Longest를 KNN에서 제거한 후, 검색된 h 번째 과거 상태벡터의 출력 상태벡터를 KNN에 추가하면서 현행 상태벡터와 가장 유사한 KNN을 구축하게 된다.

이상의 과정을 수행하기 위해서는 식(3-16)에 의해 산정된 입력변수 p 별 u_h^p 를 하나의 대표 상태거리(u_h)로 환산해주어야 한다. u_h^p 는 \mathbb{R}^p 의 차원에서 하나의 점으로 표현될 수 있으며, \mathbb{R}^p 의 차원에서 원점과 u_h^p 의 거리는 현행 상태벡터($X_{p,c}$)와의 유사성을 대표한다. \mathbb{R}^p 의 차원에서 상태거리를 산정하기에 앞서 u_h^p , $\forall p$ 는 정규화된다. 정규화를 수행하지 않을 경우, p 별 상태가 이질적임으로 스케일에 의한 차원의 저주에 빠지게 되며, 특정 p 가 전체를 대표하게 된다. 상태거리의 정규화는 KNN으로 구축된 k 개의 u_k^p 와 h 번째 u_h^p 를 포함한 집합의 최대거리와 최소거리를 이용하여 p 별로 각각 산정하며, 0.0~1.0의 값으로 정규화된다. u_k^p 와 u_h^p 의 정규화 거리를 산정하기 위하여 u_h^p 을 포함한 이웃($n=k+1$)의 상태간 거리 집

합을 $u_n^p = [u_h^p, u_1^p, u_2^p, \dots, u_k^p]$ 으로 정의한다. 이상의 정의와 식(3-17)을 이용하여 정규화된(Normalized) 상태거리($d_i^p, i \in n$)를 산정한다.

$$d_i^p = \frac{u_i^p - \min\{u_n^p\}}{\max\{u_n^p\} - \min\{u_n^p\}}, \quad \forall i, p \quad i \in n \quad \text{식(3-17)}$$

d_i^p 가 산정되면, \mathbb{R}^p 에서 i 번째 이웃에 대한 정규화된 대표 상태거리($d_i, i \in n$)는 다음의 식(3-18)을 이용하여 산정된다.

$$d_i = \|d_i\| = \left| \sum_{j=1}^p (d_i^j)^2 \right|^{1/2}, \quad \forall i, i \in n \quad \text{식(3-18)}$$

3) KNN의 DB구조

KNN 군집구축 알고리즘을 수행하기 위해서는 입·출력 상태벡터를 저장하기 위한 공간이 필요하다. 저장공간은 $k=[1, 2, 3, \dots, K, \dots, k]$ 이 이웃에 대하여 각각 ①입력자료 상태벡터 즉, 추출된 현행 상태벡터와 유사한 과거 상태벡터, ②출력 상태벡터 그리고 ③ \mathbb{R}^p 에서 대표 거리로 구성되며, <그림 3-16>과 같다. KNN DB공간은 KNN 군집구축 알고리즘의 탐색결과 저장 공간이면서 동시에 JC단계의 입력자료로 이용된다.

i	[입력 상태 벡터]	[출력 상태 벡터]	[상태 거리]
1	$[X_{v,1}, X_{b,1}, X_{q,1}]$,	$[O_{v,1}, O_{pt,1}]$,	$[u_1^p, d_1^p, d_1]$
2	$[X_{v,2}, X_{b,2}, X_{q,2}]$,	$[O_{v,2}, O_{pt,2}]$,	$[u_2^p, d_2^p, d_2]$
...
K	$[X_{v,K}, X_{b,K}, X_{q,K}]$,	$[O_{v,K}, O_{pt,K}]$,	$[u_K^p, d_K^p, d_K]$
...
k	$[X_{v,k}, X_{b,k}, X_{q,k}]$,	$[O_{v,k}, O_{pt,k}]$,	$[u_k^p, d_k^p, d_k]$

<그림 3-16> KNN 군집구축 알고리즘의 DB구조

4) KNN 군집구축 알고리즘

상태벡터, 상태간 거리 Metric, 그리고 KNN의 DB구조가 정의되면, <그림 3-17>의 KNN 군집구축 알고리즘을 이용하여 KNN DB을 구축하게 된다. 탐색된 $X_{p,h}$ 가 군집의 1차 자격을 부여받기 위해서는 식 (3-15)로 산정된 정상적인 자료의 비율(r_p , 0.0~1.0)이 수용할만한 수준의 정상 자료 비율(r_{acpt})보다 커야한다.

```

Member=0
For each [ $X_{v,h}$ ,  $X_{b,h}$ ,  $X_{q,h}$ ],  $h \in H$ 
    -Building Missing Info. ( $M^p$ ),  $\forall p$  by eq.(3-14)
    -Calculate  $r_p$ ,  $\forall p$  by eq.(3-15)
    IF  $r_{acpt} \leq r_p$ ,  $\forall p$  then
        -Calculate  $u_h^v$ ,  $u_h^b$ ,  $u_h^q$  Bet [ $X_{v,h}$ ,  $X_{b,h}$ ,  $X_{q,h}$ ] and [ $X_{v,c}$ ,  $X_{b,c}$ ,  $X_{q,c}$ ]
        by eq.(3-16)
        IF Member <  $k$ 
            -Update [ $O_{v,h}$ ,  $O_{pt,h}$ ], [ $u_h^v$ ,  $u_h^b$ ,  $u_h^q$ ]
            -Member = Member+1
        Else
            - $n = k + 1$ 
            -Calculate  $d_i^p$ ,  $\forall i, p, i \in n$  by eq.(3-17)
            -Calculate  $d_i$ ,  $\forall i, i \in n$  by eq.(3-18)
            -Search  $d_{\max}$ ,  $d_{\max} = \max\{d_1, d_2, \dots, d_n\}$ 
            -Delete [ $O_{v,j}$ ,  $O_{pt,j}$ ], [ $u_j^v$ ,  $u_j^b$ ,  $u_j^q$ ], [ $d_j^v$ ,  $d_j^b$ ,  $d_j^q$ ], [ $d_j$ ]
            from KNN DB (여기서,  $j$ 는  $d_{\max}$ 에 해당하는 자료임)
            - $n = k$ 
    End IF

```

<그림 3-17> KNN 군집구축 알고리즘

6. JC단계 모형의 개발

1) 입·출력 자료의 정의

JC단계는 KN단계에서 구축한 KNN DB의 출력값 즉, 1차 의사결정 집단을 입력값으로 군집화(Clustering) 과정을 통해 다수의 의사결정 그룹으로 분류하는 단계이다. 예측의 관점에서 JC단계는 다수의 의사결정 그룹을 결정함으로써 장래의 불확실성을 줄여주는 역할을 수행한다. 그리고 KNN-NPR 모형의 최적화 파라미터인 k 값의 지속적인 재정산이라는 단점을 극복하기 위하여 KN단계의 k 값을 크게한 후 군집화 모형을 기반으로 구축된 군집의 구성원(m)을 이용하여 다시 한번 k 값을 결정하는 역할을 수행한다. 이상의 국부적 k 값 결정과정을 통해 수용가능한 오차의 영역을 확대하는 역할을 수행한다.

JC단계의 입력값은 KN단계의 출력값을 이용하게 되며, JC단계에서 군집화를 수행하기 위한 입력값은 <그림 3-18>과 같다. JC단계의 j -군집화 모형은 \mathbb{R}^p 차원에서 k 개의 d_k^p 를 이용하여 j 개의 의사결정 그룹을 선정하며, 적정 군집(Cluster)의 개수는 적정 군집수(j_o) 결정모형을 이용하여 결정하게 된다. j -군집화 모형과 j_o 결정모형은 JC 군집화 알고리즘으로 통합된다.

i	[출력 상태벡터]	[상태거리]
1	$[O_{v,1}, O_{pt,1}]$,	$[d_1^p, d_1]$
2	$[O_{v,2}, O_{pt,2}]$,	$[d_2^p, d_2]$
...
K	$[O_{v,K}, O_{pt,K}]$,	$[d_K^p, d_K]$
...
k	$[O_{v,k}, O_{pt,k}]$,	$[d_k^p, d_k]$

<그림 3-18> JC단계 모형의 입력값

2) 기존 군집화 모형의 요구수준 적합성 분석

군집화 분석은 주어진 데이터를 직접 분석하지 않고 집단으로 정의함으로써 집단의 대표성을 찾는다는 점에서 장점이 있으며, 데이터 마이닝 분야에서 주로 이용되는 방법중 하나이다. 본 연구의 j -군집화 모형 개발에 앞서 기존 군집화 모형의 장·단점을 살펴보고, 본 연구의 요구수준 적합성을 분석함으로써 j -군집화 모형의 개발방향을 설정하였다.

다양한 군집화모형이 제시되고 있으나, 본 연구에서는 가장 대표적인 군집화 모형인 K-mean 계열에 대하여 본 연구의 요구사항에 적합한지를 검토하였다. 가장 대표적인 군집화 모형은 MacQueen(1967)과 Lloyd(1982)에 의해 개발된 K-means 모형이다. 연산속도가 빠르다는 장점이 있으나, K-means 모형은 크기가 매우 다른 군집을 찾기에는 적절하지 않으며, 이상치에 민감하고 K값을 미리 정해야 하는 단점이 있다. 이후 K-medoids가 제시되었으며(Kaufman and Rousseeuw, 1990), 이상치 등에 단점을 보완하기 위해 평균기반의 중심값(Centriod) 대신 중위값에 해당하는 Mediod를 적용한다. K-means와 K-medoids는 초기 군집화 과정의 시작위치인 중심값(Centroid) 또는 Mediod의 값에 따라 국부적 결과를 초래하는 단점이 있다. 그리고 K-mean++(Arthur and Vassilvitskii, 2007)이 제시되었으나 문제점을 완전히 해결하지는 못 하고 있다.

본 연구에서 개발되는 예측모형은 첨단자료관리시스템 등의 시스템 탑재를 목적으로 하고 있다. 따라서 적용되는 군집화 모형이 어느정도 합리적인 의사결정군집을 형성한다면, 안정적이면서 연산 수행속도가 매우 빨라야 한다. K-means의 경우, 국부 최적화에 빠질 경우 초기점을 다시 설정하고 모형을 반복 수행해야 하며, 일반적으로 100번을 반복하기 전에 수렴해야 한다. 그러나 군집의 객체수가 1개로서 분류되는 국부 최적화를 면할 수는 없다. 그리고 K-mean 계열의 모형은 타 군집화모형과 마찬가지로 시스템의 입장에서 볼 때 매우 느리다는 한계가 있다. 따라서 개발되는 군집화모형은 연산수행속도를 획기적으로 향상시키면서 이상치로 인한 국부 최적화와 초기값 설정의 문제를 극복하도록 한다.

3) j -Clustering 모형개발

기존 군집화 모형은 수행속도가 느리며, 초기점의 설정에 따라 국부 최적화 즉, 군집화 실패가 발생한다는 문제점이 있다. 예측 시스템의 관점에서 볼 때, 군집화 실패는 예측실패를 의미하며 느린 연산수행속도는 정보제공의 기회를 상실하기 때문에 예측실패와 동일한 의미이다. 따라서 JC단계에 적용되는 군집화 모형의 최우선 요구조건은 매우 빠른 연산수행속도와 더불어 국부 최적화의 사전 방지라고 할 수 있다.

본 연구에서 개발되는 j -군집화 모형은 초기점의 설정을 이용하지 않는 대신 서로간에 거리가 가장 가까운 절대 쌍(Absolute pair)로 초기화를 수행하고, 초기화된 절대 쌍을 기준으로 초기 라벨링을 수행하여 하부군집을 구축한다. 하부군집이 구축되면, 최인접 하부군집을 병합(Merging)하는 군집화 과정을 통해 안정적으로 국부 최적화를 방지하였다. 그리고 기존 군집화 모형에서 연산수행속도 저하의 원인인 라벨링을 초기 군집화 과정에서 한번만 수행하고 중심점 갱신의 반복적 과정을 생략한 최인접 군집간의 병합을 이용한 군집화 기법을 적용함으로써 연산수행속도를 극대화하였다. j -군집화 모형은 [1단계] 이상치 제거, [2단계] 최하부 군집인 절대 쌍의 탐색 및 결정, [3단계] 라벨링을 통한 초기 하부군집의 구축, [4단계] 최인접 하부군집간의 병합과정으로 구성되는 상향식 군집화 방식이다. 그리고 j -군집화 모형은 적정 군집수 결정모형과 결합된 j -군집화 알고리즘으로 구현된다.

j -군집화 모형은 KN단계의 출력값인 R^p 차원에서 k 개 요소(Element, 이하 e)를 군집화하며, d_k^p 는 R^p 에서 k -요소의 공간적 위치이다. k 개의 구성요소 집합은 $e = [e_k \mid k=1, 2, \dots, K]$ 으로 정의한다. n 개의 군집으로 구성되는 군집의 집합은 $y = [y_n \mid n=1, 2, \dots, N]$ 로 정의하며, 군집의 중심(Centriod) 집합은 $c = [c_n \mid n=1, 2, \dots, N]$ 로 정의한다.

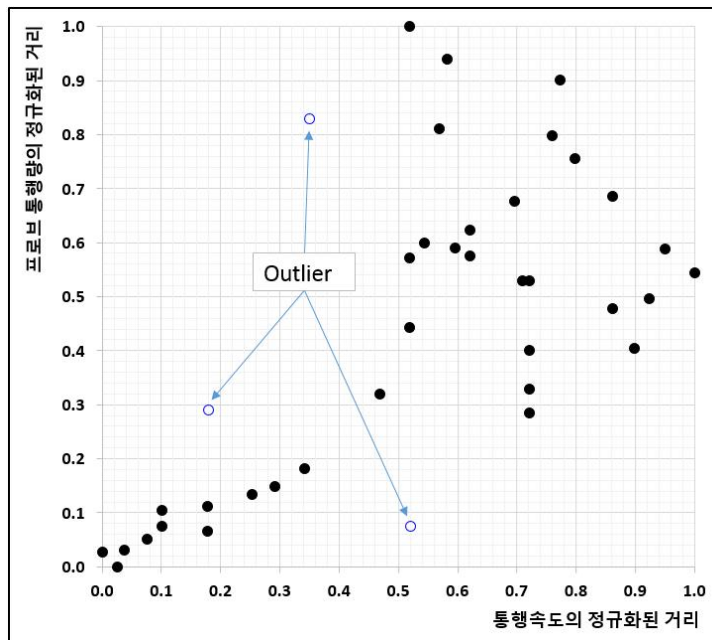
1단계의 이상치 제거는 군집화를 수행하기에 앞서 이상치로 인한 공간 왜곡을 최소화하기 위하여 수행된다. e_k 에서 최인접 요소(e_i)까지의 Euclidean 거리를 $d_{k,i}$ 로 정의하면, 식(3-19)와 같다.

$$d_{k,i} = \|e_k - e_i\|_2, \quad k \neq i, \quad i, j, k \in K \quad \text{식(3-19)}$$

식(3-19)에 의해 모든 e_k 에 대한 $d_{k,i}$ 이 산정되면, $d_{k,i}$ 집합에 대한 평균, a 와 표준편차, δ 를 산정한 후, 정상 요소의 임계 거리구간값(d_{trsh})보다 큰 경우에 해당하는 요소를 제거한다. d_{trsh} 는 식(3-20)과 같으며, t_o 는 임계값 결정 계수로서 2.0을 적용하였다.

$$d_{trsh} = a + t_o \times \delta \quad \text{식(3-20)}$$

이상치 제거의 과정은 이상치에 해당하는 모든 정보를 입력 DB에서 제외하고 k 값을 조정하는 과정으로 수행된다. <그림 3-19>는 실제 입력 자료에서 검지된 이상치의 예를 보여주고 있다. 이러한 이상치는 군집화 과정에서 군집의 중심점을 왜곡시키기 때문에 불합리한 군집화 결과를 초래할 수 있다.

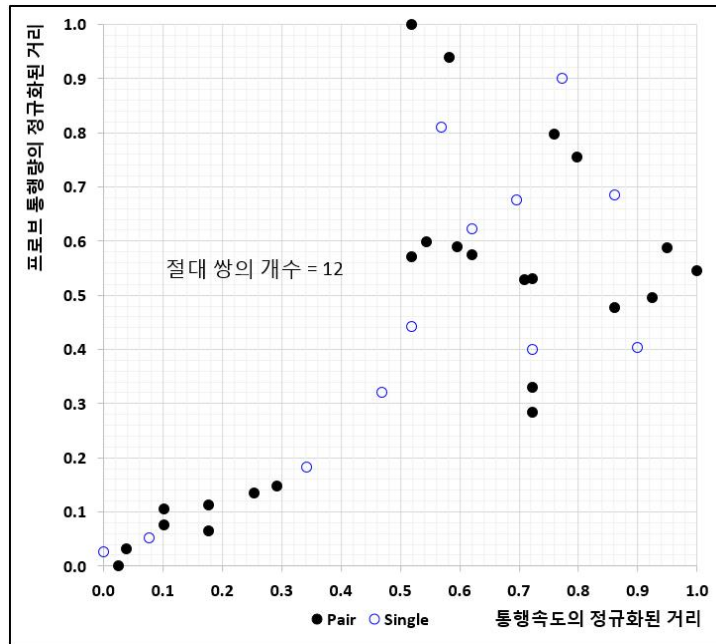


<그림 3-19> 이상치 검지 및 제거의 예

이상치가 제거된 후, 본 연구에서는 절대 쌍(Absolute pair, 이하 AP)를 결정하게 된다. 절대 쌍은 두 e_k 간의 거리가 최인점이다. 따라서 군집화 과정에서 서로 떨어질 수 없는 최하부 군집의 중심 요소라 할 수 있다. 절대 쌍 $AP(e_i, e_j)$ 의 조건은 식(3-21)과 같이 정의하며, <그림 3-20>은 구성된 $AP(e_i, e_j)$ 의 예를 보여주고 있다.

$$\arg \min \|e_i, e_k\|_2 = \arg \min \|e_j, e_k\|_2 \quad \text{식(3-21)}$$

$$i \neq j, i \neq k, j \neq k, i, j, k \in K$$



<그림 3-20> 절대 쌍 선정결과예

$AP(e_i, e_j)$ 의 개수(N_{ap})는 최하부 군집의 구성요소임으로 최대 군집의 개수로 정의할 수 있다. 즉, N_{ap} 는 군집 집합(y)의 최대 개수이다. 따라서 병합기반의 j -군집화에서는 N_{ap} 를 최대 군집의 개수(N)로 정의할 수 있다.

절대 쌍 기반의 최하부 군집화는 $AP(e_i, e_j)$ 를 e 에서 제거하고, 군집 y_n 에 $AP(e_i, e_j)$ 를 각각 추가하면서 군집의 중심인 c_n 을 식(3-22)를 이용하여 갱신한다. 초기 군집하는 모든 AP 에 대하여 수행되며, e 의 구성요소 개수는 $k = k - 2 \times N$ 으로 조정된다.

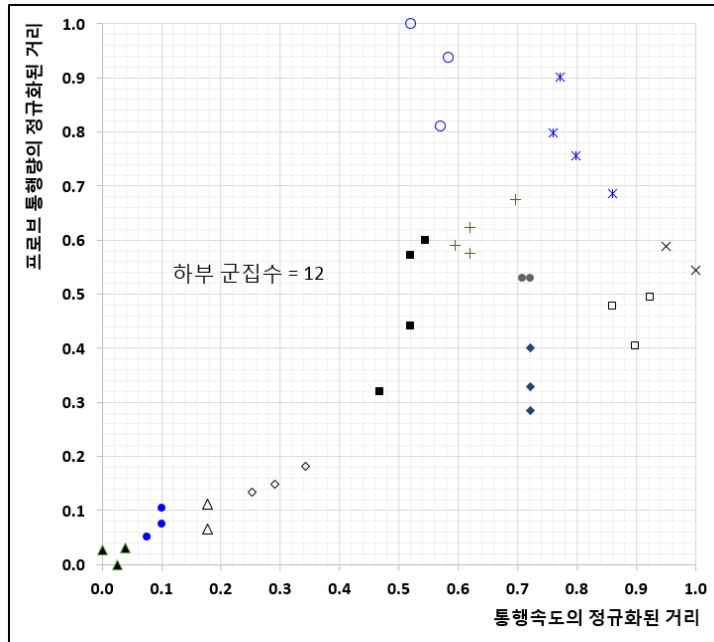
$$c_n = c(y_n) = \sum_{i=1}^{y_n} e_i / y_n \quad \text{식(3-22)}$$

최하부 군집이 구성되면, 군집에 포함되지 않은 e 의 요소에 대하여 초기 레벨링을 수행한다(참고: 레벨링은 e 의 요소를 최인접 군집에 각각 할당함을 의미함). 본 연구에서는 e_k 를 기준으로 c_n 까지 거리가 최소인 y_n 에 e_k 을 레벨링하는 기존 방식과 달리 모든 e_k 와 모든 c_n 간의 거리가 최소인 e_k 를 y_n 에 순차적으로 레벨링하게 되며, k 와 n 에 대한 조건은 식(3-23)과 같다. 선정된 e_k 를 y_n 에 레벨링한 후 식(3-22)를 이용하여 e_k 가 할당된 군집의 c_n 을 갱신한다. 그리고 $k = k - 1$ 로 조정하면서 $k = 0$ 이 될 때까지 과정을 반복한다.

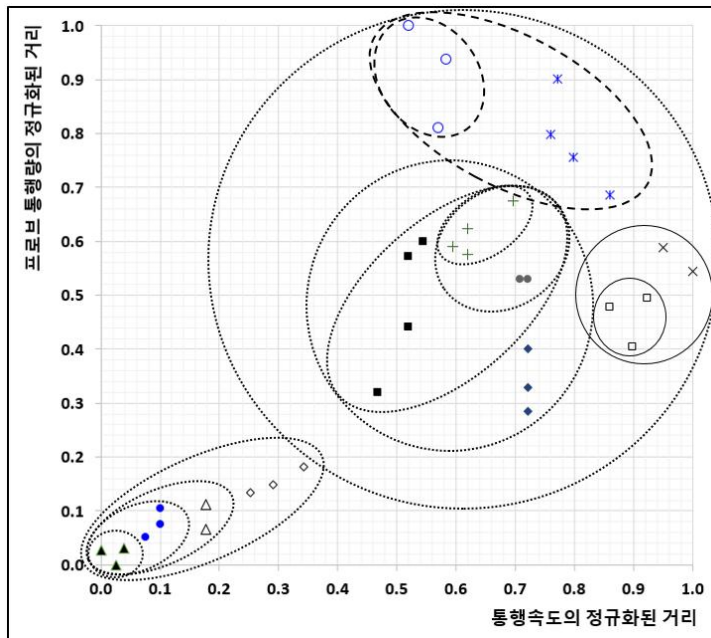
$$\arg \min \|e_k, c_n\|_2, \quad \forall k, n \quad \text{식(3-23)}$$

이상의 과정을 종료하면 <그림 3-21>과 같은 하부 군집 즉, AP기반의 최대 군집이 구축된다. 하부 군집이 구축되면 병합(Merging) 단계를 수행하게 된다. 하부 군집의 병합은 하부 군집의 중심점간 거리가 최소인 2개의 하부 군집(i, j)을 하나로 병합하는 것이며, 병합되는 하부 군집의 선정 조건은 식(3-24)와 같다. 선정된 두 군집을 병합한 후 병합된 군집의 중심점인 c_n 은 식(3-22)를 이용하여 갱신되며, $N = N - 1$ 로 조정된다. 그리고 군집 병합은 $N = 2$ 일 때까지 반복되며, <그림 3-22>는 군집병합의 과정을 예로 보여주고 있다.

$$\arg \min \|c_i, c_j\|_2, \quad \forall i, j, \quad i \neq j, \quad i, j \in N \quad \text{식(3-24)}$$



<그림 3-21> 구축된 하부 군집의 예



<그림 3-22> 군집 병합 과정의 예

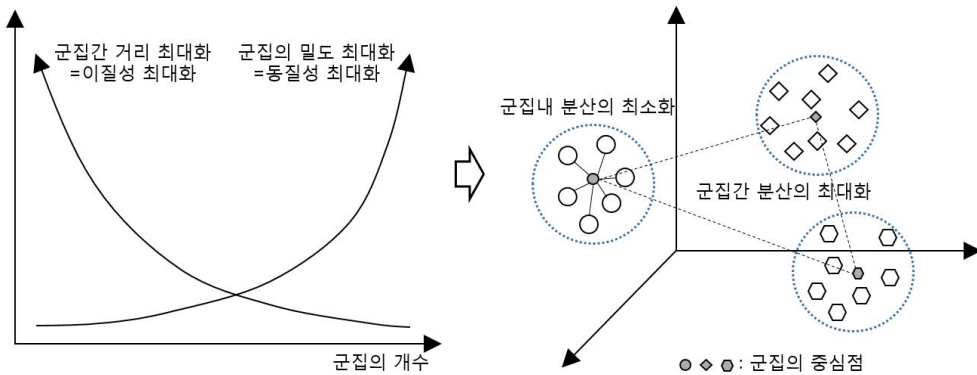
4) 적정 군집수 결정 지표

본 연구에서는 j -군집화 모형을 이용한 군집화 완료 후 적정 군집수를 결정한다. 적정 군집수는 선정된 적정 군집수 결정 모형을 이용하여 결정되며, 적정 군집수 결정모형은 j -군집화 모형과 결합된다. 군집화는 해당 군집의 동질성을 극대화함과 더불어 인접 군집과의 이질성을 극대화함으로써 패턴을 인식하는 과정이라고 할 수 있으며, <그림 3-23>과 같다. 통계학에서 군집분석인 분산분석(ANOVA)에서는 일반적으로 F-검증을 이용하며, 군집내 분산의 최소화와 군집간 분산의 최대화를 통해 군집의 패턴을 인식한다.

적정 군집수 결정에 관련한 연구⁴⁾로는 군집내 평균거리를 기반으로 하는 Elbow method가 Thorndike(1953)에 의해 제시된 후 <표 3-1>과 같은 거리제곱의 합(Sum of squares) 계열지표가 제시되었다. ANOVA 검증에서 두 집단은 F값으로 설명된다. F값은 설명되는 변량(Explained variance)과 설명되지 않는 변량(Unexplained variance)의 비율이며, 이는 집단간 변동성(Between-group variability)에 대한 집단내 변동성(Within-group variability)이다.

본 연구에서는 Clustering분야에서 적정 군집수의 결정에 주로 이용되는 F-ratio 지표를 이용하였다. F-ratio는 집단내 중심점, 즉 평균에서 집단을 구성하는 요소간의 거리제곱의 합인 SSW(Sum of squares within the clusters, 이하 SSW)에 대한 모든 요소의 중심점인 전체 평균과 집단의 중심점까지 거리제곱의 합인 SSB(Sum of squares between the clusters)의 비율이다. SSW와 SSB는 Euclidean 거리를 이용하여 식(3-25)와 식(3-26)으로 각각 산정되며, F-ratio는 SSW와 SSB를 이용하여 식(3-27)과 같이 산정된다. 여기서, N 은 구성요소의 전체 개수, k 는 군집 n 의 구성요소의 개수, e_i 는 군집 n 의 i 번째 구성요소, $c_{p(i)}$ 는 e_i 가 속한 군집의 중심점, n 은 군집의 개수, c_j 는 j 번째 군집의 중심점, \bar{x} 는 전체 군집 또는 모든 구성요소의 중심점이다.

4) 다양한 적정 군집수 결정모형이 제시되었으나, 본 연구는 가장 널리 이용되는 거리제곱의 합 모형들에 대해 언급하기로 한다.



<그림 3-23> 적정 군집수 결정의 원칙

<표 3-1> 적정 군집수 결정 모형

번호	모형	저자(연도)
1	SSW	Thorndike(1953)
2	SSW/k	Ball and Hall(1965)
3	$k^2 W $	Marriot(1971)
4	$\frac{SSB/k-1}{SSW/N-k}$	Calinski and Harabasz(1974)
5	$\log(SSB/SSW)$	Hartigan(1975)
6	$d \log(\sqrt{SSW/(dN^2)}) + \log(k)$	Xu(1997)

주1) SSW : Sum of Squares Within the clusters(MSE)

주2) SSB : Sum of Squares Between the clusters

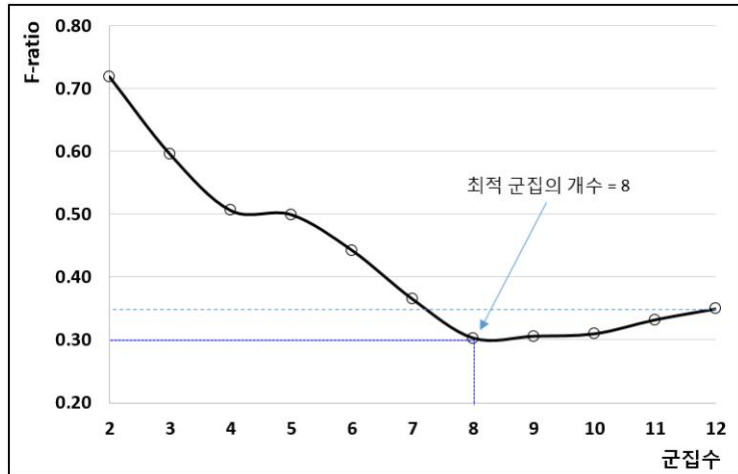
주3) d : 자료의 차원, N :자료의 크기, k : 군집의 개수

$$SSW = \sum_{i=1}^k \|e_i - c_{p(i)}\|_2 \quad \text{식(3-25)}$$

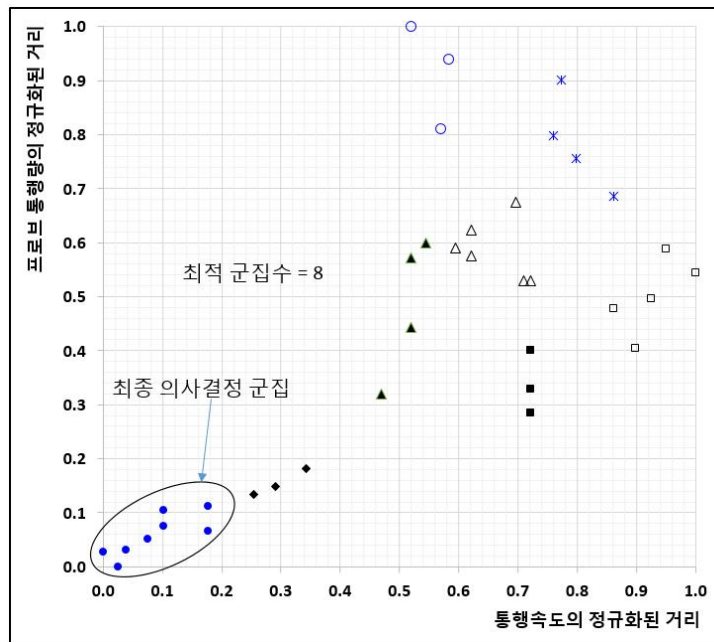
$$SSB = \sum_{j=1}^n n_j \|c_j - \bar{x}\|_2 \quad \text{식(3-26)}$$

$$F\text{-ratio} = \frac{N \cdot SSW}{SSB} = \frac{N \cdot \sum_{i=1}^k \|e_i - c_{p(i)}\|_2}{\sum_{j=1}^n n_j \|c_j - \bar{x}\|_2} \quad \text{식(3-27)}$$

적정 군집수는 F-ratio가 최소가 되는 군집수로 결정되며, <그림 3-24>는 F-ratio를 이용한 적정 군집수 결정의 예를 보여주고 있다. 그리고 <그림 3-25>는 F-ratio를 이용한 적정 군집수로 결정된 군집화 결과의 예를 보여주고 있다.



<그림 3-24> 적정 군집수 결정의 예



<그림 3-25> 최종 군집화 결과의 예

5) JC단계의 출력값

JC단계의 출력값은 입력값인 k 개의 이웃에 대한 출력 상태벡터와 상태거리와 j -군집화의 결과인 군집의 개수와 군집의 중심점에 대한 정보이며, <표 3-2>와 같다. 그리고 k 이웃 중 j -군집화 과정에서 이상치로 제거된 입력값은 제거된다. FDM단계에서 군집의 중심점(c_n)은 최종 의사결정 그룹을 결정하는데 이용되며, KN단계의 출력값인 상태거리(d_k)는 도로망 소통상태의 예측과 경로통행시간의 산정에 있어 의사결정의 가중치로 이용된다.

<표 3-2> JC단계의 출력값

i	출력 상태벡터(O)	상태거리(d_i)	군집(N)	중심좌표(c_n)
1	$O_{v,1}, O_{pt,1}$	d_1	1	c_1
2	$O_{v,2}, O_{pt,2}$	d_2		
3	$O_{v,3}, O_{pt,3}$	d_3		
4	$O_{v,4}, O_{pt,4}$	d_4	2	c_2
5	$O_{v,5}, O_{pt,5}$	d_5		
6	$O_{v,6}, O_{pt,6}$	d_6		
7	$O_{v,7}, O_{pt,7}$	d_7		
...
$k-2$	$O_{v,k-2}, O_{pt,k-2}$	d_{k-2}	n	c_n
$k-1$	$O_{v,k-1}, O_{pt,k-1}$	d_{k-1}		
k	$O_{v,k}, O_{pt,k}$	d_k		

6) j -군집화 알고리즘

j -군집화 모형과 적정 군집수 결정모형은 j -군집화 알고리즘으로 구현되며, <그림 3-26>과 같다. j -군집화 알고리즘은 이상치 제거, 절대쌍기반의 최하부 군집 구축, 레벨링을 통한 하부 군집 구축, 하부 군집의 병합, 적정 군집수의 결정으로 구성된다.

Given \mathbb{R}^p, d_k^p ,
Define $e = [e_k \mid k = 1, 2, \dots, K]$,
 $y = [y_n \mid n = 1, 2, \dots, N]$, $c = [c_n \mid n = 1, 2, \dots, N]$

STEP 1: Filtering Outlier

For all e_k
- Calculate d_k by eq.(3-19)
- Calculate d_{trsh} by eq.(3-20)

For all e_k
IF $d_{trsh} < d_k$ then
- Delete e_k from e
- $K = K - 1$

STEP 2: Finding Absolute Pair (AP)

- Find $AP(e_i, e_j)$ and N_{ap} by eq.(3-21)
- Re-define $N = N_{ap}$
- Assign AP_n onto $y_n, \forall n$
- Calculate $c_n, \forall n$
- $K = K - 2 \times N$

STEP 3: Building Sub-clusters

For all e_k
- Levelling e_k by eq.(3-23)
- Assign e_k onto the nearest y
- Update c of the nearest y

STEP 4: Merging Sub-clusters

Start Merging
- Find pair-cluster by eq.(3-24)
- Merging the pair-cluster
- Update c_n
- $N = N - 1$
IF $N = 2$ then Stop Merging
Else Find new pair-cluster

STEP 5: Deciding Suitable Number of Clusters

Decide Suitable Number of Clusters by eq.(3-27)

<그림 3-26> j -군집화 알고리즘

7. FDM단계 모형의 개발

1) 의사결정 방법론 정립

FDM단계는 <표 3-2>의 JC단계 출력값과 의사결정 규칙을 이용하여 집단의 의사결정 우선순위를 결정하고, KN단계의 출력값과 결정된 의사결정 우선순위를 이용하여 장래 도로망의 소통상태, 즉 구간별 통행속도와 경로통행시간을 예측한다. 의사결정 우선순위 결정 방법론은 과거의 사례를 기반으로 과거와 유사하거나 새로운 문제를 해결하는 사례기반 추론(Case-based reasoning, 이하 CBR)에 기초하며, CBR의 의사결정 트리(Decision making tree, 이하 DMT)는 JC단계의 출력값에 기초한다.

본 연구에서 도로망 소통상태 및 경로통행시간 예측을 위한 DMT는 의사결정 규칙에 의해 결정된다. 의사결정 우선순위는 집단간 우선순위와 집단내 의사결정 우선순위로 구분되며, 다음의 우선권 규칙(Priority rules)에 의하여 결정진다.

규칙1: 현재와 과거의 경험이 더 유사한 집단이 우선권을 가진다.

규칙2: 현재와 과거의 경험이 더 유사한 멤버가 우선권을 가진다.

이상의 규칙은 과거 경험의 유사성에 근거한다. 따라서 FDM단계의 입력값인 군집의 현재와 과거 상태의 유사성은 개별 멤버보다 우선권을 가지며, 군집 내에서 멤버들은 개별 멤버의 현재와 과거 상태의 유사성으로 우선권을 가진다. 따라서 유사성은 군집의 경우 JC단계에서 산정된 중심좌표와 원점간의 거리로, 개별 멤버의 경우 KN단계에서 산정된 상태거리로 정의한다. 이상의 정의로 의사결정 그룹과 멤버의 우선권을 결정하는 과정은 군집의 중심좌표와 원점과의 거리를 기준으로 오름차순 정렬을 수행하고 난 후, 정렬된 그룹내의 멤버에 대하여 상태간 거리를 기준으로 오름차순 정렬을 수행함으로써 의사결정권에 대한 DMT를 구성하게 된다. 그리고 결정된 DMT는 장래 상태의 결정에 적용되며, DMT의 의사결정 우선순위 결정 과정은 다음과 같다. 군집별 의사결정

권은 식(3-28)을 이용하여 R_p 차원에서 집단의 중심점과 원점간의 거리인 $d_{o,n}$ 을 Euclidean distance metric을 이용하여 산정하며, $d_{o,n}$ 이 낮은 값일 수록 과거 상태와 현행 상태가 유사함으로 우선권을 가진다.

$$d_{o,n} = \|c_n\|_2, \quad \forall n, n \in N \quad \text{식(3-28)}$$

<표 3-3>은 군집별 멤버별 의사결정 권을 부여한 DMT의 예를 보여 주고 있다. 일반적으로 의사결정권이 높은 집단 멤버의 상태간 거리는 하위 의사결정 집단 멤버들의 상태간 거리보다 낮게 나타난다. 이상의 결정된 의사결정 트리는 장래 네트워크 소통상태와 출발지-도착지간 경로통행시간 예측에 이용된다.

<표 3-3> FDM단계 의사결정권의 부여결과의 예

의사결정권 부여전					의사결정권 부여후						
N	c_n	$d_{o,n}$	d_i	출력벡터	집단 우선순위				멤버 우선순위		
					순위	N	c_n	$d_{o,n}$	순위	d_i	상태벡터
1	c_1	0.5	0.6	$O_{v,1}, O_{pt,1}$	1	3	c_3	0.1	1	0.05	$O_{v,9}, O_{pt,9}$
			0.5	$O_{v,2}, O_{pt,2}$					2	0.10	$O_{v,8}, O_{pt,8}$
			0.4	$O_{v,3}, O_{pt,3}$					3	0.15	$O_{v,7}, O_{pt,7}$
2	c_2	0.3	0.31	$O_{v,4}, O_{pt,4}$	2	2	c_2	0.3	4	0.28	$O_{v,6}, O_{pt,6}$
			0.30	$O_{v,5}, O_{pt,5}$					5	0.30	$O_{v,5}, O_{pt,5}$
			0.28	$O_{v,6}, O_{pt,6}$					6	0.31	$O_{v,4}, O_{pt,4}$
3	c_3	0.1	0.15	$O_{v,7}, O_{pt,7}$	3	1	c_1	0.5	7	0.4	$O_{v,3}, O_{pt,3}$
			0.10	$O_{v,8}, O_{pt,8}$					8	0.5	$O_{v,2}, O_{pt,2}$
			0.05	$O_{v,9}, O_{pt,9}$					9	0.6	$O_{v,1}, O_{pt,1}$
4	c_4	0.7	0.8	$O_{v,10}, O_{pt,10}$	4	4	c_4	0.7	10	0.6	$O_{v,11}, O_{pt,11}$
			0.6	$O_{v,11}, O_{pt,11}$					11	0.8	$O_{v,10}, O_{pt,10}$

2) 네트워크 소통상태 예측

네트워크 소통상태 예측은 결정된 DMT의 1순위 의사결정집단 멤버들을 이용하게 된다. 그러나 이력자료는 다량의 결측자료를 포함하고 있다. 따라서 멤버 중 장래 통행속도 자료에 결측이 발생할 경우, 하위 의사결정 집단의 멤버 중 우선권이 가장 높은 즉, 상태간 거리가 가장 짧은 멤버가 의사결정권을 부여받게 된다. 따라서 총 의사결정 멤버의 수는 1순위 의사결정집단의 멤버수(m)로 유지된다.

네트워크 소통상태 예측에 앞서 KNN-NPR의 예측식(Forecasting function, FFN)에 대하여 살펴보면 다음과 같다. FFN은 직접평균(Straight average, SA), 상태간 거리의 역수로 가중된 평균(Weighted average by the inverse of the distance, 이하 WAID) 등 다양한 FFN의 조합이 가능하다. SA는 수학적으로 상태의 중심일 뿐만 아니라 멤버들의 의사결정권을 동일하게 부여한다는 의미에서 이론적으로 일치한다. 반면, WAID는 현 상태와 유사성이 높은 과거 상태에 더 많은 장래 상태에 대한 의사결정권을 부여하는 방식이며, Heuristic한 방법이다. 그러나 다수의 연구(Smith et al., 2002; Chang et al., 2010; Chang et al., 2012_{ab}; Yoon and Chang, 2014)에서 WAID는 예측 정확도의 관점에서 SA보다 우수한 성능을 보였다. Smith et al.(2002)는 다양한 형태의 FFN을 적용하였으며, 다양한 형태의 FFN 중 SA는 하위그룹을 형성하였으며, WAID는 우수한 예측력을 보였다. 따라서 도로망 소통상태와 경로통행시간의 예측에 WAID를 적용하도록 한다.

도로망 소통상태인 구간별 통행속도의 예측은 식(3-1)의 예측문제와 동일하게 도로망을 구성하는 일련의 구간 집합, $S = \{1, 2, \dots, s\}$ 과 예측영역, $F = \{1, 2, \dots, f\}$ 의 시·공간 영역에 대하여 수행하게 된다. 도로망 소통상태는 식(3-29)와 같이 구간통행속도 출력 상태벡터(O_v)와 WAID를 이용하여 산정된다. 그리고 상태가 거리가 0일 경우 특정 멤버의 의사결정권은 무한히 증가함으로 최대 의사결정권은 $1/\epsilon$ 을 제약되며, $\epsilon = 0.001$ 를 적용하였다. 그리고 구간통행시간은 식(3-30)로 산정되며, D_s 는 구간 s , $s \in S$ 의 구간거리(km)이다.

$$LS_s = \frac{\sum_{i=1}^m O_{v,i}(i)/(d_i + \epsilon)}{\sum_{i=1}^m 1/(d_i + \epsilon)} \quad \text{식(3-29)}$$

$$LT_s = \frac{60.0 \sum_{i=1}^m D_s / (O_{v,i}(i) \times (d_i + \epsilon))}{\sum_{i=1}^m 1/(d_i + \epsilon)} \quad \text{식(3-30)}$$

3) 경로통행시간 예측

경로통행시간 예측을 위한 DMT의 멤버와 구성은 도로망 소통상태 예측과 동일하다. 그리고 경로통행시간은 동시경로통행시간(Instantaneous path travel time, 이하 IPTT), 경험기반 경로통행시간(Experience-based path travel time, 이하 EPTT), 동적 구간기반 경로통행시간(Dynamic link-based path travel time, 이하 DLPTT), 경로기반 경로통행시간(Path-based path travel time, 이하 PPTT)의 4가지 경로통행시간을 예측하게 된다.

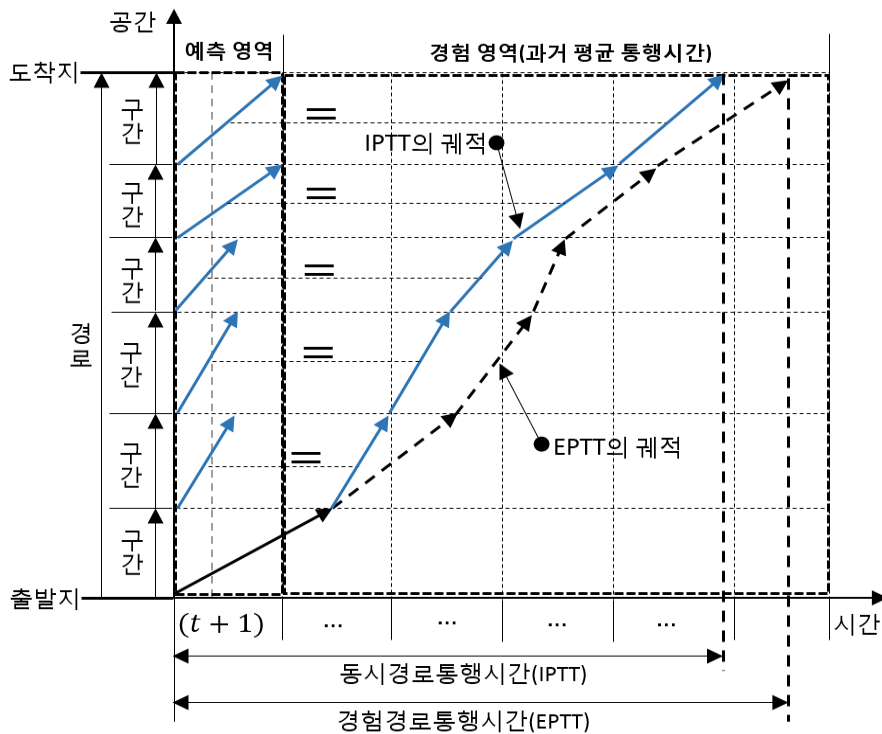
<그림 3-27>은 IPTT와 EPTT의 차이를 예측영역과 차량의 이동궤적으로 보여주고 있다. IPTT는 예측된 시점 $t+1$ 의 구간통행시간이 장래에도 지속적으로 유지된다는 가정하에 경로통행시간을 산출한다. 그리고 EPTT는 시점 $t+1$ 이후의 장래 구간통행시간은 과거의 패턴을 따른다고 가정하며, 두 경로통행시간의 예측식은 다음과 같다.

IPTT는 식(3-30)으로 예측된 시점 $t+1$ 에서 구간통행시간이 장래에도 동일하다는 조건을 부여함으로 시점 $t+1$ 에서 o,d 간 경로상에 존재하는 모든 구간의 통행시간 합으로 표현이 가능하다. 따라서 o,d 간 IPTT는 식(3-31)과 같이 직접적으로 산출이 가능하다. 여기서, n 은 o,d 간에 존재하는 연속된 구간의 개수이다.

$$IPTT_{o,d}(t+1) = \sum_{i=1}^n LT_i(t+1) \quad \text{식(3-31)}$$

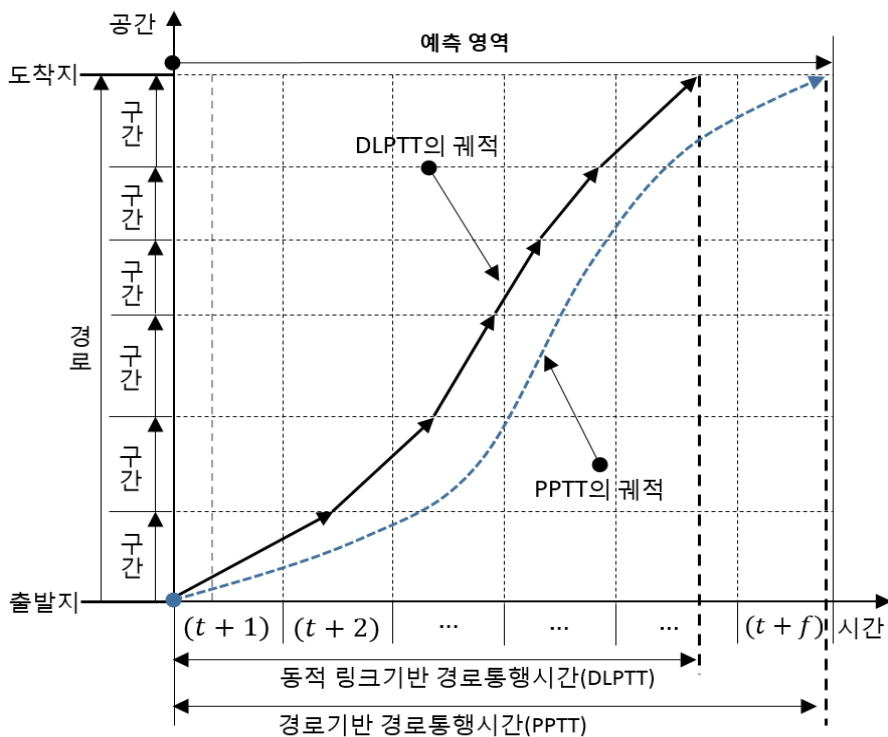
EPTT는 예측된 시점 $t+1$ 이후의 장래상태는 과거 통행시간 패턴을 따른다는 조건이 부여됨으로 IPTT와 같은 직접 산출방식을 적용할 수 없다. EPTT의 산정은 경로상에 존재하는 구간의 진입시점과 진입시점에서 과거 통행시간을 이용하여 구간의 진출시점을 연속적으로 산출하면서 목적지까지 구간통행시간을 누적하게 되며, 예측식은 식(3-32)와 같다. 여기서, n 은 $o \rightarrow d$ 간 경로에 존재하는 공간적으로 연속된 구간의 집합, t_{sum}^i 는 $o \rightarrow d$ 간 경로에 존재하는 $i(i \in n)$ 번째 구간의 진입시점으로서 $i-1$ 번째 구간의 진출시점까지의 누적 구간통행시간, $LT_i(t_{sum}^i)$ 은 i 번째 구간의 진입 시점(t_{sum}^i)에서 구간통행시간이다. 그리고 시점 $t+1$ 이상의 경우 Historical speed profile을 이용해 산정된 구간통행시간을 적용한다.

$$EPTT_{o,d}(t+1) = \sum_{i=1}^n LT_i(t_{sum}^i) \quad \text{식(3-31)}$$



<그림 3-27> IPTT와 EPTT의 예측영역과 경로통행시간 궤적

IPTT와 EPTT의 산정을 위한 장래 예측영역(Forecasting horizon)은 단기 통행시간 예측모형의 예측영역으로 한정된다. 반면, DLPTT의 예측영역은 차량의 목적지 도착 이후까지 확장되며, PPTT의 예측영역은 무한대로 확장하는 동적 경로통행시간의 개념으로서 <그림 3-28>과 같다. DLPTT는 기존 EPTT의 경험영역인 과거 평균 구간통행시간대신 식(3-30)으로 예측된 장래 6시간까지의 구간통행시간을 이용하여 동적 경로통행시간을 산출한다. PPTT는 장래 구간통행시간의 예측없이 과거의 개별 경로통행시간 자료를 이용하여 경로통행시간을 직접 예측하는 방법으로서 대표 구간통행시간에서 발생하는 오차를 감소시킴과 동시에 예측영역을 무한히 확장하는 것과 같다. 그리고 PPTT는 단일 경로가 아닌 S집합의 영역에 존재하는 모든 존간 $o-d$ 쌍에 대한 경로통행시간을 예측한다.



<그림 3-28> DLPTT와 PPTT의 예측영역과 경로통행시간 궤적

DLPTT 예측식은 식(3-33)과 같으며, EPTT에서 이용하는 Historical speed profile로 산정된 구간통행시간을 적용하지 않고 식(3-30)으로 예측된 도로망의 구간통행시간을 이용한다. 여기서, n 은 $o \rightarrow d$ 간 경로에 존재하는 연속된 도로구간의 집합, t_{sum}^i 은 $o \rightarrow d$ 간 경로에 존재하는 $i(i \in n)$ 번째 구간의 진입시점으로서 $i-1$ 번째 구간의 진출시점까지의 누적 구간통행시간, $LT_i(t_{sum}^i)$ 은 i 번째 구간의 진입 시점(t_{sum}^i)에서 구간통행시간이다.

$$DLPTT_{o,d}(t+1) = \sum_{i=1}^n LT_i(t_{sum}^i) \quad \text{식(3-33)}$$

PPTT는 출력 상태벡터 $O_{pt,m}$ 을 이용하여 시점 $t+1$ 에서 출발하는 차량의 모든 o, d 쌍간 경로통행시간을 예측하며, 식(3-30)의 구간통행시간 예측식과 마찬가지로 WAID를 이용하여 식(3-34)와 같이 산정한다.

$$PPTT_{o,d}(t+1) = \frac{\sum_{i=1}^m O_{pt,i}(i)/(d_i + \epsilon)}{\sum_{i=1}^m 1/(d_i + \epsilon)} \quad \text{식(3-34)}$$

4) FDM 알고리즘

JC단계의 출력값을 입력값으로 이용하는 FDM 단계는 FDM 알고리즘으로 구현되며, <그림 3-29>와 같다. FDM 알고리즘은 군집 우선순위와 군집내 멤버의 우선순위를 이용한 CBR기반 의사결정트리, 의사결정트리를 이용한 네트워크 장애 소통상태 예측, 예측된 네트워크 소통상태와 4개 경로통행시간 예측식을 이용한 경로통행시간의 예측으로 구성된다.

Given $[c_n, d_{o,n}]$, $[O_{v,k}, O_{pt,k}, d_k]$, $n \in N$, $k \in K$

STEP 1: Assigning Priorities based on $d_{o,n}$ and d_k

-Arrange $[c_n, d_{o,n}]$ in ascending order of $d_{o,n}$ by Priority Rule1
(Priority Rule1: the priority rule of cluster)

For all c_n

-Arrange $[O_{v,k}, O_{pt,k}, d_k]$ in ascending order of d_k by Priority Rule2
(Priority Rule2: the priority rule of within-cluster member)

STEP 2: Forecasting Future traffic condition

-Estimate future network traffic states for all the TIs by eq.(3-29)
-Estimate future link travel times for all the TIs by eq.(3-30)

STEP 3: Estimating Path travel time

For all OD pairs

-Calculate $IPTT_{o,d}(t+1)$ by eq.(3-31)
-Calculate $EPTT_{o,d}(t+1)$ by eq.(3-32)
-Calculate $DLPTT_{o,d}(t+1)$ by eq.(3-33)
-Calculate $PPTT_{o,d}(t+1)$ by eq.(3-34)

<그림 3-29> FDM 알고리즘

8. 통합 KJC 예측기의 수행구조

KN, JC, FDM 단계의 KNN 구축 알고리즘, j -Clustering 알고리즘, FDM 알고리즘은 <그림 3-30>과 같이 KJC 예측기로 통합된다. KJC 예측기는 모듈형 구조로서 모듈간 입·출력값의 구조를 이용한 구조변경이 용이하다. 그리고 모듈간 입·출력값은 단일-단일, 단일-다중, 다중-단일, 다중-다중으로 정의가 가능하며, 모듈별 고려사항 및 기능에 대하여 설명하면 다음과 같다.

KNN 구축 알고리즘의 주된 고려사항은 장래 상태의 불확실성 최소화, On-line 기능, 검색속도 최대화, 누락자료의 처리 그리고 입·출력자료의

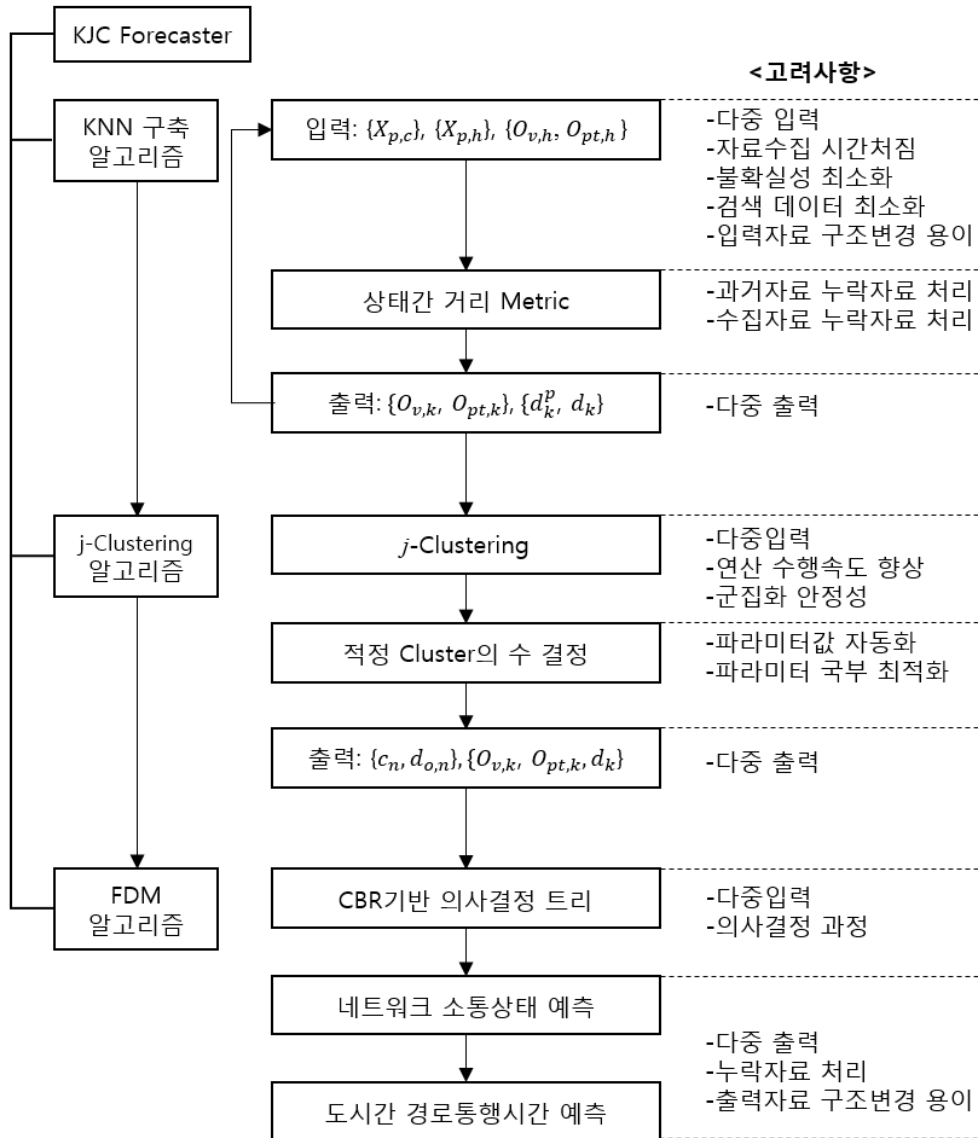
구조변경 용이성이다. 첫째, 일반적인 구간통행시간 예측모형은 단일 또는 여러 개의 구간으로 구성되는 공간을 정의한 후, 공간에서 시계열 상태변화를 이용하여 장래 상태의 방향성과 변화량을 추정함에 있어 불확실성을 최소화하고자 한다. 그러나 개발모형은 장래 상태의 불확실성을 최소화하기 위한 방안으로 네트워크 단위의 입력자료를 이용함으로써 국부적인 즉 일개 구간의 불확실성 감소보다는 네트워크의 전반적인 소통 상태를 이용하여 불확실성을 감소시킨다. 둘째, 구간검지체계에서 실시간 자료의 수집에 발생하는 시간처짐을 고려한 입력자료를 이용함으로써 On-line 적용을 고려하였다. 따라서 KNN 구축 알고리즘은 지점 및 구간 검지체계에서 수집되는 모든 자료를 이용할 수 있을 뿐만 아니라 Off-line으로 적용이 가능하다. 셋째, 과거 상태벡터의 정의에 있어 시간적 종속성을 부여함으로써 검색되는 이력자료의 양을 최소화하였다. 단기예측의 경우 시간종속성을 적용하지 않을 경우 보다 많은 다양성을 확보하면서 예측성능을 향상시킬 수 있지만 이력자료의 검색에 상당한 시간이 소요된다. 반면, 현행 상태의 국부적 변동은 중·장기 상태변화에 대한 설명력을 저하시킨다. 따라서 시간적 종속성을 적용하는 방안은 검색자료의 양을 감소시키면서 장래 상태의 예측력을 향상시키는 방안 중 하나이다(Chang et al., 2012_b). 넷째, 누락자료의 모형내 처리이다. 일반적으로 방대한 양의 이력자료와 실시간 수집자료에는 다양한 형태의 누락자료가 포함되어 있다. 자료 보정(Imputation)과 같은 누락자료를 추정하는 다양한 방법론들이 제시되고 있으나, 개발 모형을 적용하기 위하여 또 다른 하위 시스템을 개발하는 것은 개발 모형의 적용성을 현저히 감소시키는 실질적인 원인이 된다. 따라서 상태간 거리의 산정시 누락자료의 처리와 더불어 상태간 거리를 조정함으로써 누락자료에 대한 문제를 해결하였다. 마지막으로 KNN은 설명변수와 종속변수간의 관계를 고려하지 않으며 어떤 통계적 분포도 가정하지 않는다. 따라서 입·출력자료의 구조변경이 용이할 뿐만 아니라 새로운 입력자료와 출력자료의 추가가 용이하다.

j -Clustering 알고리즘의 역할은 파라미터 자동정산 및 연산수행속도 향상이다. 첫째, 기존 KNN 예측방법론의 전역 최소화 파라미터인 현행 상태 시계열자료의 개수(l)과 k 값은 월단위 주단위 또는 일단위로 조정되어야 한다. l 값은 한번 최적화를 통해 고정값으로 적용이 가능한 반면, k 값은 지속적인 조정이 필요하다. 따라서 k 값은 KNN기반 예측모형의 예측력에 영향을 미치게 된다. 추가로 k 값 정산에 있어 교통상태에 대한 국부적 값이 아닌 전일에 대한 전역 최소화를 이용할 경우 예측오차는 증가한다. j -Clustering 알고리즘은 KNN 구축 알고리즘으로 구성된 k 개의 이웃에 대한 국부 군집을 구축한 후 이를 국부 k 값으로 이용함으로써 모형의 연산과정에서 k 값을 최적화한다. 둘째, 군집화 모형의 느린 연산수행속도는 시스템의 요구시간을 충족하지 못하는 경우가 발생한다. 따라서 j -Clustering 알고리즘은 절대 쌍을 기반으로 하는 하부 군집과 하부 군집의 병합방법을 이용하여 연산수행속도를 최소화하였다. 마지막으로 기존 군집화 방법에서 발생하는 군집의 객체가 일개인 상태 즉, 군집화 실패를 절대 쌍을 이용하여 사전에 방지함으로써 시스템 적용시 안정성을 확보하였다.

FDM 알고리즘은 장래 도로망의 소통상태 예측과 경로통행시간 예측 기능을 수행하며, 다음을 고려하였다. 첫째, 군집과 군집내 멤버의 장래 상태 결정에 있어 우선권을 이용한 CBR기반의 의사결정나무(Decision making tree)를 이용함으로써 합리적 의사결정과정을 도모하였다. 둘째, 장래 소통상태 및 경로통행시간의 예측에 있어 이용되는 이력자료 즉, 출력 상태벡터에 결측자료가 발생할 경우, 의사결정나무의 우선권을 이용하여 결측자료를 보충함으로써 예측에 있어 안정성을 고려하였다. 셋째, 네트워크 단위로 장래 소통상태를 예측함으로써 구간단위 예측모형에 비하여 모형의 연산수행속도를 획기적으로 향상시켰다. 마지막으로 경로통행시간의 예측은 동적 구간기반 경로통행시간(DLPTT)과 경로기반 경로통행시간(PPTT)을 선택할 수 있게 함으로써 다중경로와 단일경로의 다양한 환경에서 적용성을 고려하였다.

추가로 KJC 예측기의 모듈형 수행구조는 예측기의 알고리즘 변경 및

구조변경에 장점을 가진다. 현행 상태벡터와 과거 입력 상태벡터의 유사성에 Kennel 거리(Kennel et a., 1992)나 쌍체 T-검정 등을 이용할 수 있다. 그리고 Clustering 단계의 경우, 예측대상의 특성과 목적에 따라 고도화된 군집화 모형과 최적 군집개수 결정모형을 적용할 수 있다.



<그림 3-30> 통합 KJC 예측기의 수행구조

IV. 자료구축 방법론

1. 개요

1) 배경 및 목적

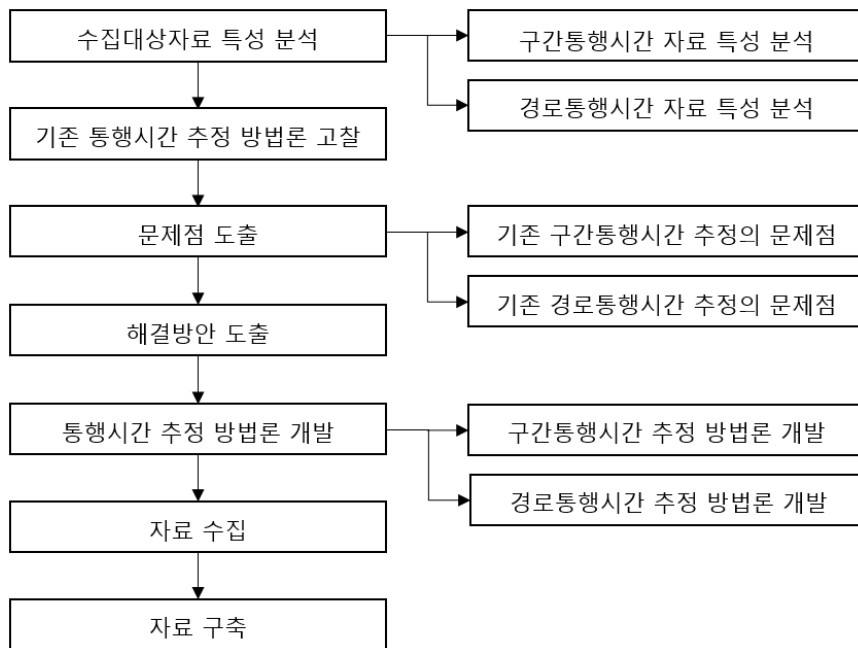
3장에서 설명된 KJC 예측기는 대용량 자료를 이용하기 위하여 KNN-NPR를 기반으로 j -군집화모형과 결합되어 개발되었다. KNN-NPR 예측모형의 정확도는 이력자료의 개수(n) $\rightarrow\infty$ 이고 KNN의 $k\rightarrow\infty$ 인 조건에서 $k/n\rightarrow 0$ 인 상태에 따라 좌우된다. 다시말해, 개발모형의 예측성능은 이력자료의 양과 질(Quantity and quality) 그리고 자료에 내재되어 있는 정보의 다양성(Variety)에 전적으로 의존한다. 따라서 개발모형의 성능평가에 앞서, 개발모형의 입력자료 요구조건인 양과 질이 우선적으로 충족되어야 한다.

개발모형의 1차 목표는 장래 6시간까지 도로망의 소통상태를 예측하는 것이며, 2차 목표는 예측된 도로망 소통상태 정보를 이용하여 중·장거리 경로통행시간을 예측하는 것이다. KJC 예측기의 성능평가를 위한 자료는 3장에서 기술한 개발모형의 입·출력 자료에 적합하도록 구축하였다. 구축자료는 한국도로공사의 근거리전용통신(DSRC) 기술에 기반을 둔 구간검지체계에서 수집·집계되는 구간통행시간, 구간 프로브 통행량, IC/TG간 경로통행시간과 한국도로공사의 전자요금징수시스템인 TCS에서 수집·집계되는 영업소 유·출입 통행량이다. 그리고 구축자료의 핵심인 이력자료의 신뢰성을 담보하기 위하여 구간 및 경로 통행시간 추정 방법론을 개발하여 자료구축에 적용한다.

추가로 본 연구에서 개발된 KJC 예측기는 입·출력 변수의 구조변경이 용이하도록 설계되었으므로 시계열 형식의 교통자료이면 입·출력 변수로 정의가 가능하다. 따라서 개발모형의 일반 도로망 적용시 AVI, GPS 등의 기술을 이용하는 구간검지체계에서 수집되는 구간통행시간 또는 속도, 구간 프로브 통행량 자료를 적용해도 된다.

2) 자료구축 방법론의 구성

자료구축 방법론은 <그림 4-1>과 같이 순차적으로 구성된다. 수집 대상자료인 구간통행시간과 경로통행시간의 특성을 분석하여 통행시간 추정 방법론의 요구조건을 검토하였다. 분석된 구간 및 경로 통행시간 자료의 특성과 통행시간 추정 방법론의 요구조건을 고려하여 기존 통행시간 추정 방법론을 고찰한 후 기존 방법론의 문제점을 구간통행시간과 경로통행시간 추정으로 구분하여 도출하였다. 도출된 문제점의 해결방안을 마련하고, 이를 기반으로 구간통행시간과 경로통행시간 추정 방법론과 알고리즘을 각각 개발하였다. 자료구축을 위한 선행 작업으로 자료구축의 방향을 설정한 후, 수집된 자료와 개발된 구간 및 경로 통행시간 추정 방법론을 이용하여 개발된 예측모형의 평가를 위한 자료를 구축하였다. 마지막으로 구축된 자료의 특성을 사례를 통해 예측의 관점에서 설명하도록 한다.

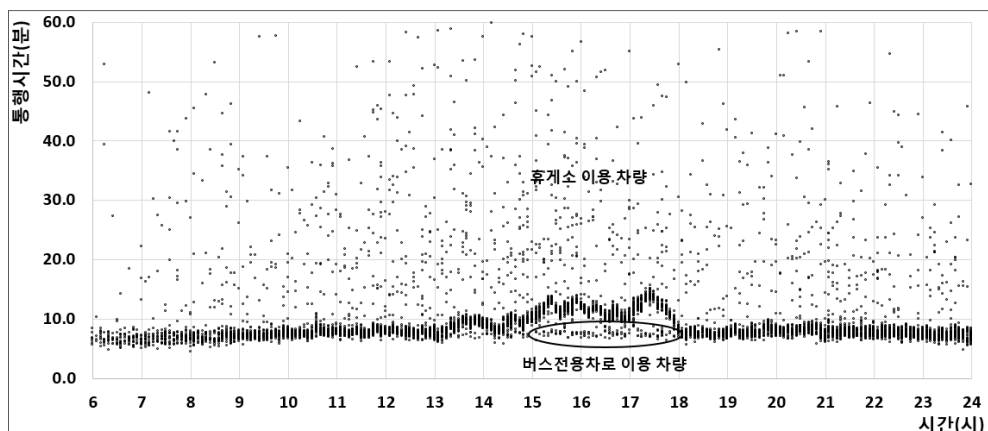


<그림 4-1> 자료구축 방법론의 구성

2. 수집대상자료의 특성분석

1) 구간통행시간 자료의 특성분석

일반적으로 구간검지체계에서 구간통행시간은 t 시점에 해당구간에 진입하여 시점 $t + \alpha_i$ 에 구간을 진출한 개별 차량(i)들의 구간통행시간(α_i)에 대한 대푯값을 의미한다. 다시말해 주어진 시간대 동안 수집된 구간통행시간 그룹에 대한 대푯값을 의미한다. <그림 4-2>는 경부고속도로 상행 안성휴게소 구간을 6~24시 사이에 통행한 1종(일반 승용차)와 6종(소형 승용차) 차량의 개별 구간통행시간을 보여주고 있으며, 통행시간 그룹은 명확하게 구분되고 있다. 첫 번째 행태는 휴게소를 이용한 차량과 비이용 차량으로 구분되고 있다. 두 번째 행태는 혼잡시 휴게소 이용 차량, 휴게소를 이용하지 않은 일반차로 이용차량 그리고 버스전용차로 이용차량으로 구분되고 있으며, 비 혼잡시의 경우 버스전용차로가 운영중이더라도 첫 번째 행태를 보이고 있다. 따라서 구간통행시간의 대푯값은 휴게소를 이용하지 않고 (혼잡시) 버스전용차로를 이용하지 않은 차량의 구간통행시간을 이용하여 추정해야 할 것으로 판단된다.

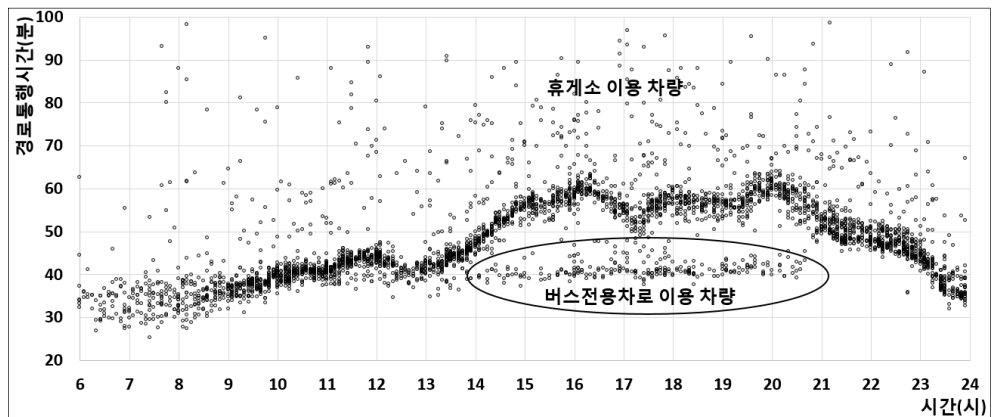


<그림 4-2> 개별 구간통행시간 자료의 특성

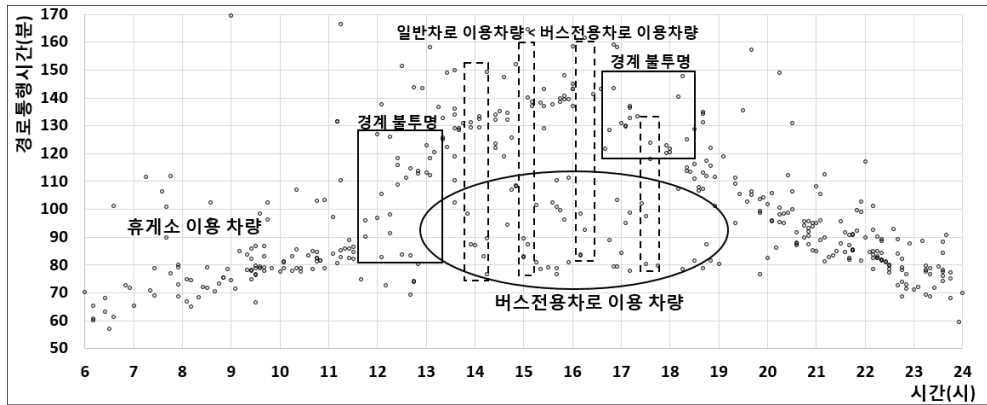
2) 경로통행시간 자료의 특성분석

경로통행시간은 시간대(t) 동안 특정 지점을 출발하여 $t+\alpha_i$ 에 특정 지점에 도착한 개별 차량들(i)의 총 통행시간(α_i)에 대한 대푯값을 의미한다. <그림 4-3>의 천안IC→서울TG(63.2km)간 일반 승용차(1종, 6종)기준 개별 경로통행시간 행태는 <그림 4-2>의 구간통행시간 행태와 유사하게 나타나고 있으며, 경로통행시간의 대푯값을 산정할 수 있을 정도로 충분한 프로브 대수를 보이고 있다.

반면, <그림 4-4>의 대전IC→서울TG(130.9 Km)간 장거리 개별 경로통행시간의 행태는 대푯값을 산정하기에 충분하지 않게 나타나고 있다. 버스전용차로 이용차량과 일반차로 이용차량의 경계가 불분명하게 나타나고 있으며, 일반차로 이용차량과 휴게소 이용차량의 경계 또한 불분명하다. 이는 통행시간이 증가함에 따라 휴게소 이용시간이 짧은 차량의 통행시간과 낮은 주행속도 그룹에 속한 차량의 통행시간이 특정 통행시간 영역에서 혼재하기 때문이다. 그리고 버스전용차로 이용차량의 표본수가 일반차로 이용차량의 표본수보다 크게 나타나는 경우가 발생하고 있다. 추가적으로 일반차로 이용 차량의 통행시간 표본이 부족하기 때문에 통행시간의 증가 및 감소 상태에서 경계가 불투명하게 나타나고 있다.



<그림 4-3> 개별 경로통행시간 자료의 특성: 중거리 통행



<그림 4-4> 개별 경로통행시간 자료의 특성: 장거리 통행

3) 통행시간 추정 방법론의 요구조건 분석

구간 및 경로 통행시간 개별 자료의 특성분석 결과는 다음과 같다. 첫째, 구간통행시간과 프로브 표본수가 충분히 수집되는 경로통행시간은 버스전용차로의 운영 및 혼잡의 발생 여부에 따라 2~3개의 그룹으로 구분되었다. 둘째, 장거리 경로통행시간의 경우 프로브 표본수가 현저히 감소하기 때문에 통행시간 그룹간의 경계가 불분명하며, 통행시간의 증가 및 감소 단계에서 또한 경계가 불분명하다. 그리고 임의적 변동을 수반하는 적은 프로브 표본수로 구성되는 경로통행시간 그룹이 전체 경로통행시간을 대표하는 경우가 발생하며, 이는 통행시간의 과대/과소 추정문제를 발생시킨다.

따라서 통행시간 추정모형은 다음의 요구조건을 충족해야 한다. 첫째, (버스전용차로 등의 도로운영 조건 변화, 통행시간의 증가, 그리고 휴게소 이용차량 등으로 형성되는) 다수의 통행시간 그룹 중 일반차로를 이용하면서 휴게소를 이용하지 않은 대표 통행시간 그룹을 선정하고, 선정된 통행시간 그룹의 개별 통행시간 자료를 이용하여 대표 통행시간을 추정해야 한다. 둘째, 프로브 표본수가 부족한 상태 즉, 통행시간 그룹이 불분명한 조건에서도 신뢰할 수 있는 대표 통행시간을 추정해야 한다.

3. 기존 방법론 고찰

1) 기존 통행시간 추정 방법론 고찰

구간검지체계에서 신뢰성 높은 구간 또는 경로 통행시간 정보의 생성은 수집자료의 정확도와 더불어 신뢰성 높은 대푯값을 추정⁵⁾하는 문제로 정의할 수 있다. 이러한 통행시간 대푯값 추정의 문제는 이상치 제거 과정과 대푯값 산정으로 구성된다.

대표 통행시간의 추정에 가장 널리 이용되는 기법은 중위절대편차(Median absolute deviation, 이하 MAD)를 이용한 이상치 제거이다. MAD기법은 표준편차 대신에 MAD를 이용하여 이상치를 제거하므로 자료의 분포를 가정하지 않아도 된다. 반면 MAD기법은 자료의 표본수가 낮거나 분산이 큰 경우 대푯값을 왜곡하는 문제가 발생한다. 다시말해 대푯값의 과대/과소 추정문제를 피할 수는 없다. MAD기법은 식(4-1)과 같으며, 여기서 tt_i 는 i 번째 통행시간 자료(tt), tt_{med} 는 tt 계열의 중위값, z_i^{MAD} 는 MAD에 의한 표준점수(Z score), 그리고 1.4826은 MAD를 정규 분포의 표준편차와 동일하게 만드는 계수이다.

$$MAD = 1.4826 \times \text{median}|tt_i - tt_{med}| \quad \text{식(4-1)}$$
$$z_i^{MAD} = \frac{tt_i - tt_{med}}{MAD}$$

Transmit 알고리즘(Mouskos et al., 1998)은 AVI을 통해 수집되는 구간통행시간 자료의 통행시간 추정방법으로서 뉴욕과 뉴저지의 교통관리 시스템에서 적용되고 있다. Transmit 알고리즘은 이상치를 제거하지 않고 15분 수집주기 평활화 기법을 이용하며, k 시간대의 A-B구간에서 관측된 통행시간의 평균통행시간(tt_{ABk})은 식(4-2)와 같이 산정된다. 여기서

5) 본 연구에서 추정(Estimation)은 시점 t 를 기준으로 과거 교통자료에 대한 대푯값의 산출을 의미하며, 예측(Forecasting)은 시점 t 를 기준으로 미래의 교통자료에 대한 대푯값의 산출을 의미한다.

t_{Ai} 는 A지점에서 i 번째 차량이 검지된 시각, t_{Bi} 는 B지점에서 i 번째 차량이 검지된 시각, n_k 는 k 시간대에 관측된 프로브의 대수이다.

$$tt_{ABk} = \frac{\sum_{i=1}^{n_k} (t_{Bi} - t_{Ai})}{n_k} \quad \text{식(4-2)}$$

해당 수집주기 동안의 구간통행시간을 식(4-2)를 이용하여 산정한 후, 과거의 (동요일, 동시간대) 패턴자료를 이용해 평활화한 다음 이전 주기 동안 평활화된 통행시간을 이용하여 현재 주기의 평활화된 대표 구간통행시간을 추정한다. 그리고 평활화된 통행시간은 식(4-3)과 같이 산정된다. 여기서 tth''_{ABk} 는 k 시간대의 평활화된 구간통행시간, tth''_{ABk-1} 는 $k-1$ 시간대의 평활화된 구간통행시간, tth_{ABk} 는 k 시간대의 평활화된 과거 구간통행시간, 그리고 α 는 평활화 계수이다.

$$tth''_{ABk} = \alpha \times tth_{ABk} + (1 - \alpha) \times tth''_{ABk-1} \quad \text{식(4-3)}$$

Transmit 알고리즘의 평활화 계수에 대한 근거는 불확실하다. α 는 0.1을 적용하며, 교통사고의 발생시 0.0을 적용한다. 따라서 교통사고 발생시 교통사고 발생시 평활화된 과거 구간통행시간을 이용하지 않으며, 증분된 자료수집의 시간처집 만큼의 이전 시간대 통행시간을 이용하게 된다. Transmit 알고리즘은 구간통행시간의 비반복적 특성이 강하고 대표 구간통행시간과 차이가 큰 이상치가 발생하는 경우 평활화 기법만으로 통행시간을 산출하기에는 한계가 있다.

Ohba et al.(1999)는 고속도로 TCS 자료의 이상치를 제거하는 방법론과 적은 표본수에서 결측치 보정에 관한 방법을 제시하였다. 이상치 제거 방법은 수집된 통행시간의 빈도수를 이용하거나 군집분석을 수행하였으나, 명확한 방법론은 제시되지 않았다. 단일 구간의 정상 교통류 상태에서 수집된 자료에 적용하였으며, 여러 개의 연속된 구간이나 정체시

적용에는 한계가 있다. 적은 표본수의 경우, 비 혼잡시의 평균을 이용하였으나 구체적인 방안을 제시하지 않았다.

강진기 등(2002)은 통계적으로 상한과 하한을 초과한 값을 제거한 후 신뢰구간 68.26%(평균 $\pm\sigma$) 이내의 자료를 이용하는 신뢰구간 추출법을 제시하였으며, 대표 구간통행시간은 식(4-4)와 같이 산정된다. 여기서, \bar{tt} 는 구간통행시간의 산술평균, tt_i 는 수집된 개별 구간통행시간, σ 는 수집된 개별 구간통행시간의 표준편차, n 은 상한값과 하한값을 제외한 표본수, tt_{is} 는 수집된 개별 구간통행시간 자료 중 신뢰구간 내에 존재하는 자료이다. 신뢰구간 추출법은 수집된 표본의 정규성을 가정하기 때문에 과분산이 발생할 경우 통행시간의 범위가 커지게 된다. 따라서 대푯값인 평균이 과대 산정되는 문제가 발생하며, 표본수가 적으면 대표 통행시간의 기능을 상실할 수 있다.

$$\begin{aligned} \bar{tt} &= \sum tt_{is} / n \\ tt_{is} &= |tt_i - \bar{tt}| \leq \sigma \end{aligned} \quad \text{식(4-4)}$$

TransGuide 알고리즘은 미국 샌 안토니오의 고속도로 교통관리시스템인 TransGuide에서 운영 중인 구간통행시간 추정 알고리즘으로써 Dion and Rakha(2003)에 의해 개발되었다. TransGuide 알고리즘은 식(4-5)를 이용하여 사전에 입력된 경계값을 초과하는 경우 이상치를 제거한 후 이동평균 기법인 식(4-6)을 이용하여 구간통행시간을 추정한다. 여기서 t 는 프로브 수집주기, Stt_{ABi} 은 시간대 t 동안 A-B구간에서 수집된 유효 프로브 표본수, t_{Ai} 은 차량 i 의 A지점 통과시각, t_{Bi} 은 차량 i 의 B지점 통과시각, t_w 는 이동평균의 시간대 길이(Rolling average window), l_{th} 는 구간통행시간 파라미터(0.2), tt_{ABi} 는 t 동안 A-B구간에서 수집된 유효 표본의 평균통행시간, 그리고 tt'_{ABi} 은 주기 $t-1$ 동안 A-B구간에서 수집된 유효 표본의 평균통행시간이다.

$$Stt_{ABi} = \left\{ t_{Bi} - t_{Ai} \mid t - t_w < t_{Bi} < t \text{ and } tt_{ABi}(1 - l_{th}) \leq t_{Bi} - t_{Ai} \leq tt_{ABi}(1 + l_{th}) \right\} \quad \text{식(4-5)}$$

$$tt_{ABi} = \frac{\sum_{i=1}^{|Stt_{ABi}|} (t_{Bi} - t_{Ai})}{|Stt_{ABi}|} \quad \text{식(4-6)}$$

TransGuide 알고리즘의 성능은 l_{th} 와 t_w 에 의해 좌우된다. l_{th} 는 이전 주기와 현재 주기 즉, $t-1$ 시간대와 t 시간대의 유효 통행시간 차이이며 0.2를 이용한다. 따라서 t 시간대에 수집된 표본 구간통행시간이 $t-1$ 시간대의 구간통행시간 추정값과 20%이상 차이가 나면 이상치로 판단하여 제거한다. t_w 는 보통 2분을 적용하며, $t-t_w$ 와 t 동안에 구간의 하류부 즉 B지점을 통과한 차량들의 통행시간을 유효 표본수로 이용한다. TransGuide 알고리즘은 l_{th} 가 상수값으로 설정되어 있으므로 혼잡의 상태나 구간거리에 따른 통행시간의 변동량을 고려할 수 없다. 다시 말해 구간통행시간이 급격히 증가하는 경우 이를 이상치로 판단하게 되며, 최초 수집된 표본에 편향이 발생할 경우 정상치가 제거되고 이상치가 대푯값이 될 수 있기 때문에 실제 구간통행시간을 왜곡시킬 수 있다.

이후 Dion and Rakha(2006)은 TransGuide 알고리즘과 Transmit 알고리즘의 단점을 보완한 알고리즘을 제시하였다. 개선된 알고리즘은 자유교통류 상태에서 통행시간은 정규분포를 따르고, 혼잡교통류 상태에서 log를 취한 통행시간은 정규분포를 따른다는 사실을 이용한다. $t-1$ 시간대까지 평활화된 통행시간과 평활화된 표준편차를 이용하여 t 시간대의 유효 구간통행시간 표본을 결정하고, 식(4-6)과 같이 산술평균하여 대표 구간통행시간을 추정한다. 그러나 자유교통류 상태의 매우 낮은 밀도에서 이질적 구간통행시간을 설명하는데 한계가 있으며, 동시 교통류 상태에서 구간통행시간의 변동을 고려하지 않는다. 따라서 기존의 TransGuide 알고리즘과 Transmit 알고리즘의 성능은 개선되었으나, 앞에서 언급한 두 알고리즘의 근본적 문제점을 개선하지는 못 하였다.

남궁성(2005)은 인접 TG간 경로통행시간을 이용하여 중·장거리 경로 통행시간을 추정하는 전진반복전후방탐색법(PIFAB)을 제시하였다. PIFAB 모형은 TCS 통행시간에 포함되어 있는 휴게소 체류시간 등 고속도로 본선 외의 통행시간을 제거하면서 낮은 프로브 표본수를 보이는 중·장거리 경로통행시간을 추정할 수 있다는 장점이 있다. 반면, 경로내에 TG간 표본수가 충분히 확보되어야 하며, 결측구간이 발생하였을 경우 연속성을 상실할 가능성이 있다.

Oliver and Nour-Eddin(2006)은 상·하류부 구간 ETC자료를 이용하여 TG간 단거리 구간통행시간을 추정하였다. 3개의 결절점으로 구성된 2개의 연속적인 구간 A-B-C에서 A-B구간의 구간통행시간 표본수가 부족한 경우, A-C구간의 통행시간에서 B-C구간의 통행시간을 차감하여 A-B구간의 통행시간을 산정하였다. 이러한 구간통행시간 차감 방법은 수집되는 통행시간 표본이 적은 경우 유용하게 적용될 수 있지만, 자료 수집에 발생하는 시간처짐을 극복하지 못하는 한계가 있다. 그리고 단거리 구간의 통행시간의 추정에는 유효하지만, 장거리 통행이 높은 지역간 도로에는 적용이 어렵다는 한계를 가지고 있다.

2) 기존 방법론의 문제점 분석

기존 통행시간 추정 방법론에 대한 고찰결과를 정리하면 다음과 같다. 첫째, 기존의 통행시간 추정 방법론은 남궁성(2005)의 연구를 제외하면 구간통행시간 추정에 목표를 두고 있으며, 단거리 구간의 통행시간의 추정에 적합하도록 개발되었다. 따라서 상대적으로 구간의 길이가 긴 경우나 경로통행시간의 추정에는 적합하지 않은 것으로 판단된다.

둘째, 기존 통행시간추정 방법론이 개발될 당시의 구간정보수집기술과 수집되는 통행시간 표본수는 낮은 실정이었다. 따라서 대표 통행시간 그룹을 제외한 나머지는 이상치로 간주하였다. 통행시간 분포에 관한 일반적인 견해는 “통행시간 분포는 정규분포를 따르며, 이를 벗어난 이상치

가 존재한다”이다. 이러한 통행시간 분포에 대한 일반적인 견해는 기존 통행시간 추정 방법론에 대한 이론적 근거가 되어왔다. 그러나 현재의 구간검지기술은 과거에 비해 비약적으로 발전하였다. <그림 4-2>와 <그림 4-3>에서 살펴본 바와 같이, 통행시간 구간검지를 위한 스마트 단말의 급속한 보급으로 단일 구간과 단·중거리 경로통행의 경우 프로부 수집율은 매우 높은 상태이며, 다수의 통행시간 그룹이 존재하였었다. 따라서 기존의 통행시간 추정방법론의 1차 단계인 이상치 제거에 대한 접근법은 실제 적용에 있어 불합리한 결과를 초래할 수 있다. 이는 수집되는 통행시간의 분포가 다수의 그룹으로 구성됨으로 정규분포가 아닌 경우가 대다수 발생하며, 과분산을 피할 수 없기 때문이다.

셋째, 많은 연구에서 이상치 제거를 위해 Box-Plot법, Shapiro-Wilk 검정 등의 통계적 방법을 적용하였다. 이러한 통계적 방법들은 분포의 정규성을 가정하기 때문에 분포의 중심에서 벗어난 통행시간 자료는 이상치로 간주한다. 그러나 구간 및 경로 통행시간 분포는 좌편향과 과분산의 특성을 보이는 경우가 많다. 따라서 기존의 통계적 방법을 이상치 제거에 이용하기에는 한계가 있다.

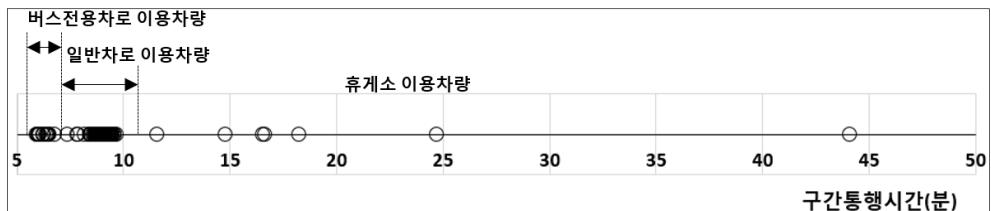
넷째, Transmit 알고리즘의 평활화 계수($\alpha=0.1$)과 TransGuide 알고리즘의 이상치 제거 파라미터($l_{th}=0.2$)는 상수로서 현행 시간대에서 교통류 상태를 고려할 수 없다. 즉, 구간통행시간의 변동량에 대한 고려가 미흡하기 때문에 통행시간이 급격히 변화하는 경우 긴 상태회복(State recovery) 시간을 피할 수 없다.

마지막으로 오늘날의 발전된 구간검지체계에서도 장거리 경로통행시간의 표본수는 <그림 4-4>에서와 같이 적을 뿐만 아니라 통행시간의 경계상태가 불분명하다. 이는 통행거리 빈도분포의 특성상 통행거리가 증가함에 따라 통행량 즉, 수집되는 경로통행시간 표본수가 급격히 감소하기 때문이다. 따라서 장거리 통행시간의 추정에 기존 통행시간 추정 방법론을 적용할 경우 극단적인 과대/과소 평가를 피할 수 없다.

3) 해결방안 도출

기존 구간통행시간 추정 방법론의 문제점은 낮은 표본 수집 상태에서 하나의 대표 통행시간 그룹을 가정하고 나머지 통행시간은 이상치로 간주하는 이상치 제거에 있었다. 그러나 최근의 발단된 구간검지기술과 높은 표본수는 기존의 통행시간 추정 기법과 다른 방법론이 필요함을 의미한다.

<그림 4-5>는 경부고속도로 상행선 안성 휴게소구간에서 수집된 승용차⁶⁾의 개별 프로브 구간통행시간 분포를 보여주고 있으며, 5분 수집주기 동안 104건의 수집되었다. 기존의 낮은 표본수 문제는 해결되었지만, 기존의 이상치는 표본수가 증가하면서 명확한 구간통행시간 그룹을 형성하고 있다. 분포의 평균(μ)은 9.75분, 중위값(m)은 8.94, 표준편차(σ)는 5.88로 분석되었다. 상·하한값을 $m \pm 2\sigma$ 로 설정할 경우 유효 통행시간 범위는 -2.82~20.7로서 버스전용차로 이용차량과 휴게소 이용차량의 구간통행시간을 포함하여 대표 통행시간을 추정하게 된다. 따라서 휴게소 이용 그룹이 없는 경우 과소추정, 휴게소 그룹이 있는 경우 과대추정의 문제가 발생하며, 엄밀히 말해 대표 구간통행시간 그룹의 대푯값이라 하기 어렵다. 이상의 과대/과소 추정문제를 해결하기 위한 방안 중 하나는 버스전용차로와 휴게소를 이용하지 않은 일반차로 이용차량을 대표 그룹으로 선정한 후 대표 구간통행시간을 산정하는 것이다. 그리고 대표 그룹은 군집분석과 같은 방법으로 결정할 수 있다.



<그림 4-5> 개별 구간통행시간의 분포

구

6) 본 연구에서 차종은 한국도로공사의 6개 요금징수 차종 중 1종과 6종인 승용차와 경차를 대상으로 한다.

구간통행시간 그룹은 버스전용차로 이용그룹, 일반차로 이용그룹(고속, 중속, 저속), 그리고 휴게소 이용그룹으로 구분되며, 교통상태에 따라 <표 4-1>과 같이 4개 그룹군으로 구분할 수 있다. 4개 그룹군 중 대표 그룹은 가장 높은 빈도수를 보이는 그룹으로 선정해야 한다. 따라서 비혼잡시 중속도 그룹, 혼잡시 저속도 그룹이 대표 구간통행시간 그룹으로 선정되는 것이 합리적이다.

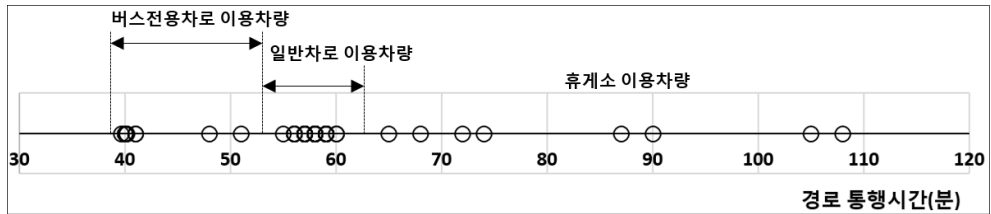
<표 4-1> 구간통행속도 그룹군 분류

구분		그룹군	구간통행속도 그룹			버스전용차로 이용 그룹	휴게소 이용 그룹
			고속	중속	저속		
교통 상태	비혼잡	1	○	○	○	-	○
		2	○	○	○	-	-
	혼잡	3	-	-	○	○	○
		4	-	-	○	○	-

남궁성(2005)의 연구를 제외하면 대부분의 기존 방법론은 구간통행시간의 추정을 목표로 하고 있었으며, 중·장거리 경로통행시간의 추정에는 적합하지 않은 것으로 분석되었다. <그림 4-6>에서 단·중거리 개별 경로통행시간 분포의 특성은 절대적인 통행시간의 증가량을 제외하면 개별 구간통행시간의 특성과 유사하며, <표 4-1>의 통행시간 그룹군 3과 동일하게 나타나고 있다.

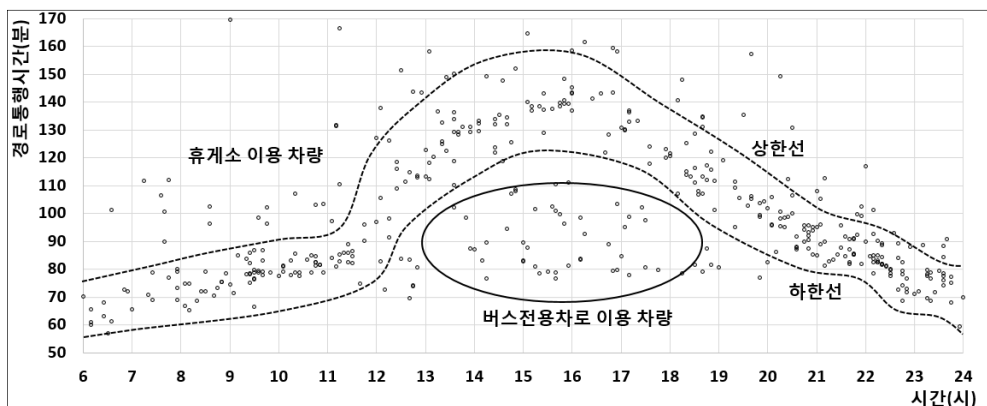
<그림 4-6>은 경부고속도로 상행선 천안IC→서울TG간의 개별 프로브 경로통행시간 분포를 보여주고 있으며, 표본수는 55(건/5분)이다. 경로통행시간 표본임에도 표본수는 높게 나타나 대푯값을 산정하는데 무리가 없는 것으로 판단되며, 구간통행시간 그룹과 동일하게 3개 그룹을 형성하고 있다. 분포의 중위값(m)은 58.0, 표준편차(σ)는 13.98로서 상·하한값을 $m \pm 2\sigma$ 로 설정할 경우 유효 통행시간 범위는 30.04~85.96이다. 상·하한값으로 유효 경로통행시간을 선정할 경우 자유교통류 상태에서 과속차량과 혼잡시 일반차량의 경로통행시간을 모두 포함하게 된다. 따라서 기존의 통계적 방법을 이용하여 이상치를 제거하고 대표 경로통행

시간을 추정하는 경우 과대/과소추정의 문제를 피하기 어려울 것으로 판단된다. 이상의 과대/과소 추정문제를 해결하기 위한 방안 중 하나는 구간통행시간의 방안과 마찬가지로 군집분석 등의 기법을 이용하여 일반차로 이용차량의 경로통행시간 그룹을 대표 그룹으로 선정하는 것이다.



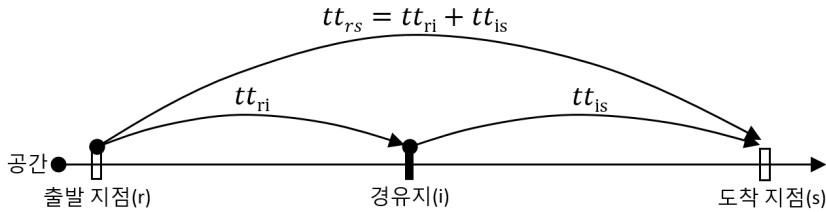
<그림 4-6> 개별 경로통행시간의 분포

장거리 경로통행시간의 경우 적은 표본수로 인하여 대표 경로통행시간 그룹의 선정시 문제가 발생한다. 이는 통행거리가 증가함에 따라 프로브 대수는 급감하기 때문이다. <그림 4-7>은 대전IC→서울TG 경로통행시간 분포를 시간대별로 보여주고 있다. 버스전용차로 이용차량과 휴게소 이용차량이 일반차로 이용차량과 혼재되어 있기 때문에 일반차로 이용차량 그룹의 경계가 명확하지 않으며, 통행시간 상·하한선이 불분명하다. 따라서 군집분석 등의 기법을 적용한다 하더라도 상·하한선의 영역내에서 경로통행시간의 과대/과소 추정문제를 해결하기 어렵다.



<그림 4-7> 장거리 경로통행시간의 모호한 경계 특성

반면, 단·중거리 경로통행시간은 <그림 4-3>과 같이 일반차로 이용 차량 그룹의 경계가 명확하기 때문에 <그림 4-6>과 같이 집계 시간대 별로 대표 그룹을 선정할 수 있다. 따라서 장거리 경로통행시간은 <그림 4-8>과 같이 경유지를 기반으로 하는 일종의 경로통행시간 릴레이 기법을 이용하여 출발지에서 도착지까지 경로통행시간 산정이 가능하다. 이와 유사한 접근법으로 남궁성(2005)은 인접 TG간 TCS 통행시간 자료를 이용하여 경로통행시간 수집률이 낮은 고속도로 장거리 경로통행시간을 추정하였으며, 합리적인 결과를 도출하였다.



<그림 4-8> 경유지기반 릴레이 경로통행시간 산출의 개념

출발지(r)에서 시각 t_o 에 출발한 차량은 경유지(i)까지 $tt_{r,i}(t_o)$ 의 통행시간이 소요되며, 다시 경유지에서 시각 $t_o + tt_{r,i}(t_o)$ 에 출발한 차량은 도착지(s)까지 $tt_{i,s}(t_o + tt_{r,i}(t_o))$ 의 통행시간이 소요된다. 따라서 시각 t_o 에서 출발한 차량의 $r \rightarrow s$ 경로통행시간, $tt_{r,s}(t_o)$ 은 식(4-7)과 같이 산정이 가능하다. 그리고 경유지는 차량 단말기(OBU)가 통과하는 노변기지국(RSE)으로 설정이 가능하다. 이상에서 설명한 경유지기반 릴레이 경로통행시간 산출개념은 순차적으로 RSE를 추가할 경우 통행거리를 무한히 연장할 수 있다. 또한 단일 RSE가 아닌 인접 RSE를 그룹으로 설정할 경우 일부 RSE의 장애가 발생하는 자료수집 실패의 상태에서도 장거리 경로통행시간을 산출할 수 있는 장점이 있다.

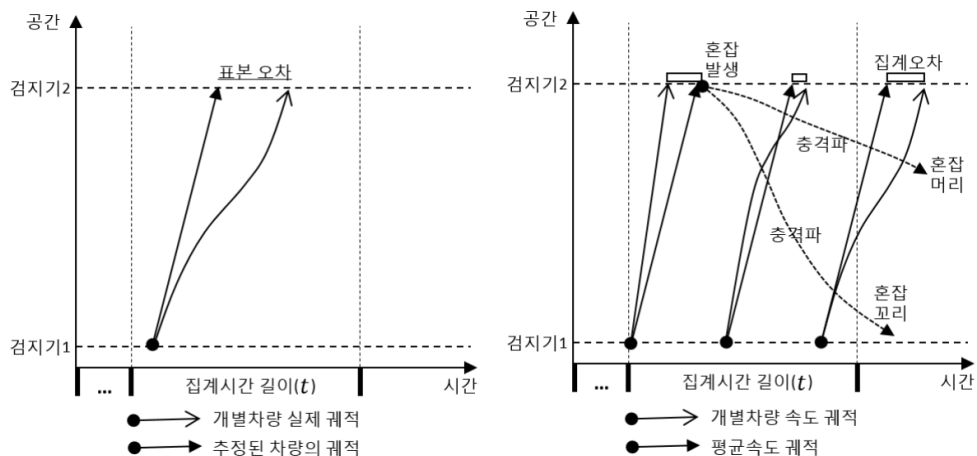
$$tt_{r,s}(t_o) = tt_{r,i}(t_o) + tt_{i,s}(t_o + tt_{r,i}(t_o)) \quad \text{식(4-7)}$$

4. 통행시간 추정 방법론 개발

1) 집계시간 길이와 최소 표본수의 결정

구간검지체계에서 자료는 개별 구간통행시간으로 수집되며, 집계시간 길이의 수집주기 동안 수집된 자료는 이상치 제거와 같은 전처리 과정을 거친 후 집계된다. 따라서 구간 및 경로 통행시간 자료를 특정 방법을 이용하여 집계하기 위해서는 집계시간 길이의 설정이 선행되어야 한다. 그리고 집계된 자료의 신뢰성을 담보하기 위해서는 최소 표본수를 만족해야 한다. 그러므로 통행시간 추정 방법론의 개발에 앞서 기존 연구의 결과를 기반으로 집계시간의 길이와 최소 표본수를 설정하도록 한다.

수집자료의 집계과정에서 발생하는 통행시간 오차는 표본오차, 그룹오차, 그리고 산술평균 오차로 구분할 수 있으며, <그림 4-9>와 같다. 표본오차는 수집된 표본으로 추정된 차량의 주행궤적과 실제 주행궤적의 차이에서 발생한다. 집계오차는 집계된 추정 통행시간과 수집된 개별 통행시간간의 오차이다. 산술평균 오차는 산술평균이 조화평균보다 크거나 같기 때문에 산술평균으로 산정된 그룹의 통행시간은 조화평균으로 산정된 그룹의 통행시간보다 과소 평가되는 오차이다. 그리고 이러한 오차는 집계시간의 길이, 표본수 등과 밀접한 관련이 있다.



<그림 4-9> 통행시간 추정오차의 종류

Smith and Ulmer(2003)은 통계적 검증기법을 이용하여 교통량의 집계 시간 길이에 관한 연구를 수행하였으며, 15분이내에서 가장 길게 설정하는 것이 바람직하다고 지적하였다. Liu et al.(2008)은 첨두 시간대의 경우 정보 손실을 최소화하기 위하여 집계시간 길이는 짧게 해야하며, 비첨두 시간대의 경우 집계시간대 길이를 길게해야 한다고 지적하였다. Wang and Liu (2005)와 Bigazzi et al.(2010)은 집계시간의 길이를 연장하면 대표 통행 시간의 추정오차는 감소하며, 집계시간의 길이가 5분을 초과하면 다시 추정오차는 증가함을 보였다. 따라서 적정 통행시간 집계시간의 길이는 5분이라고 지적하였다. 다시 말해 집계시간 길이를 5분 보다 짧게하거나 길게할수록 추정된 통행시간의 대표성과 신뢰도는 감소함을 의미한다.

교통정보제공 주기는 일반적으로 집계시간 길이의 n 배수로 결정되지만, 정보의 짧은 생애주기 특성과 신속한 정보제공을 위하여 동일하게 적용하고 있다. 추가로 대부분의 국내 ITS시스템은 집계시간의 길이와 교통정보 제공주기로 5분을 적용하고 있다. 이상의 이론과 실제 시스템의 운영상태를 고려하여 본 연구의 수집자료 집계시간 길이와 교통정보 제공주기는 5분으로 설정하였다. 따라서 개발모형의 시간 단위인 시간대의 길이(TL)는 5분으로 설정된다.

집계시간의 길이가 정의되면 집계된 통행시간 자료의 대표성과 신뢰성을 확보하기 위한 수집주기 동안의 최소 표본수와 대푯값 산정 방법을 결정해야 한다. 최소 표본수(n_{\min})는 통행시간이 정규분포라는 가정하에 통계적 기법을 이용하여 결정하거나, 표본 분포의 가정없이 충분한 표본수가 수집된다는 가정하에 중심극한정리를 적용하여 식(4-8)과 같이 산정된다(Chen and Chien, 2000; 이영인과 이정희, 2002). 여기서, σ_{kt} 는 이력자료의 표준편차, μ_{kt} 는 이력자료의 평균, ϵ_{\max} 는 최대 상대오차, r 은 절대 상대오차가 ϵ_{\max} 보다 작은 시간대의 비율이다.

$$n_{\min} = \left[\left\{ \Phi^{-1} \left(\frac{1+r}{2} \right) \left(\frac{\sigma_{kt}}{\mu_{kt}} \right) \right\} / \epsilon_{\max} \right]^2 \quad \text{식(4-8)}$$

중심극한정리에 의한 최소표본수(n_{\min})는 모든 구간과 시간대에 동일한 변동계수(r)를 적용하기 때문에 교통류의 동적 특성을 고려하지 못한다. 그리고 실제 적용시 구간통행시간이나 경로통행시간의 표본 수집율은 n_{\min} 을 충족하기 어려운 경우가 다수 발생하기 때문에 중심극한 정리의 적용에는 한계가 있다. 심상우 등(2013)은 서울-대전간 경로통행시간의 경우 중심극한 정리를 적용할 경우 5분 집락에서 13.4대의 최소 표본수가 요구된다고 지적하였다. 추가로 최소 표본수는 이상치가 제거된 대표 통행시간 그룹의 최소 표본수라는 점에서 실제 최소 표본수는 n_{\min} 보다 커야 한다. 따라서 본 연구의 경우 구간 및 경로 통행시간의 대푯값 추정을 위한 최소 표본수는 20건/5분으로 설정하였다.

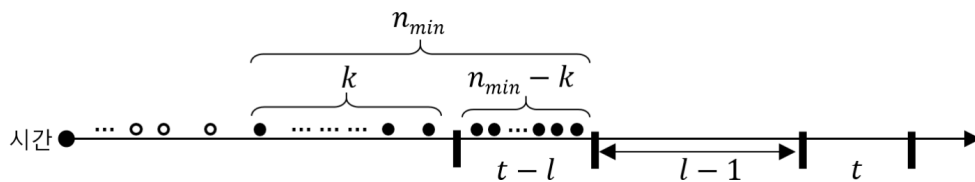
이상치 제거 과정을 통해 선별된 최종 통행시간 그룹에 대한 대푯값으로 산술평균, 조화평균, 중위값 등이 적용되고 있다. 지점검지기반의 경우 산술평균에 의한 오차를 제외하기 위하여 조화평균을 이용하여 통행속도를 산정한다. 그러나 통행시간의 경우 조화평균을 이용하여 대푯값을 산정하게 되면 과대 추정의 문제가 발생한다(Coifman et al., 2003). 추가로 통행시간은 좌편향 분포를 보이는 경우가 많기 때문에 평균은 적합하지 않다. Coifman et al.(2003)은 중위값을 이용할 경우 분포를 가정하지 않아도 되며, 산술평균이나 조화평균의 과대/과소 추정 문제를 어느정도 극복하면서 신뢰성이 높은 대푯값으로 적용할 수 있음을 보였다. 따라서 본 연구에서는 선별된 개별 통행시간으로 구성되는 통행시간 그룹의 대푯값으로 중위값을 이용하도록 한다.

2) 구간통행시간 추정 방법론 개발

구간통행시간 추정방법론은 적용방식에 따라 On-line과 Off-line모형으로 구분된다. On-line 통행시간 추정은 구간검지체계하에서 자료의 수집에 발생하는 시간차집 상태에서 가장 최근에 수집된 자료의 통행시간을 산출하는 방식이다. Off-line 추정은 기 수집된 과거자료를 이용하여 해당 집계시간대에 대한 통행시간을 산출하는 방식으로써 자료수집에 발

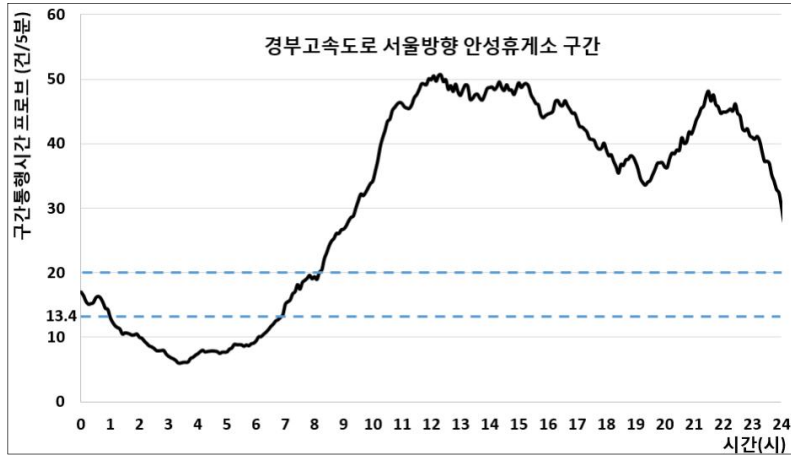
생하는 시간처짐을 고려하지 않는다. 추가로 3장에서 서술한 장애 소통 상태 예측모형은 KN단계에서 통행속도 자료의 수집에 발생하는 시간처짐이 고려함으로써 On-line 적용을 목적으로 개발되었다.

본 연구에서 개발된 구간통행시간 추정모형은 <그림 4-5>와 같이 충분한 구간통행시간 표본수가 수집되는 상태에서 최소 표본수를 만족시킨 후 대표 통행시간 그룹을 선별하게 되며, 대표 그룹의 선별과정에서 타 통행시간 그룹과 이상치를 점진적으로 동시에 제거하며, 다음과 같다. 개별 구간통행시간 자료수집에 발생하는 시간처짐(l) 상태에서 집계 시간대 길이(TL)로 구분되는 수집주기($t-l$)에 수집된 개별 구간통행시간 집합을 $x = [x_n \mid n=1, 2, \dots, N]$ 로 정의한다. 그리고 $N < n_{min}$ 인 경우 <그림 4-10>과 같이 집계 시간대 $t-l$ 에서 시간적으로 가까운 $k(=n_{min}-N)$ 개의 이전 수집주기 구간통행시간 표본을 x 에 추가한다.



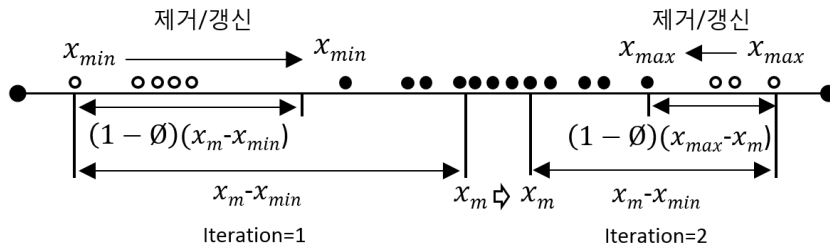
<그림 4-10> 이전 시간대 방향 집계시간 확장

이전 수집주기의 개별 자료를 추가로 이용하는 것은 Transmit 알고리즘의 평활화(α)와 TransGuide 알고리즘의 이동창(l_{th})과 유사한 개념이지만, 고정값이 아닌 표본수의 부족시 표본수의 크기에 따라 동적으로 조정된다는 점에서 다르다. 그리고 이전 수집주기의 자료를 이용하는 경우는 <그림 4-11>과 같이 표본수가 낮은 심야~아침 시간대이며, 대부분의 시간대에서 최소 표본수를 만족한다. 이러한 비침두시 집락 시간대 길이의 연장은 Liu et al.(2008)이 지적한 바와 같이 추정된 구간통행시간이나 통행속도의 신뢰도를 향상시키게 된다. 이상의 과정을 통해 x 가 구축되면, $N \geq n_{min}$ 인 조건에서 최소 표본수를 만족하게 된다.



<그림 4-11> 시간대별 구간통행시간 프로브 변동

최소 표본수를 만족하는 x 가 구축되면, 대표 구간통행시간 그룹의 선정을 위하여 타 구간통행시간 그룹과 이상치를 점진적으로 제거하게 된다. 다시 말해 대표 구간통행시간 그룹을 구성한다고 추정되는 멤버들을 제외한 나머지를 제거하는 방법이라고 할 수 있다. 대표 구간통행시간 그룹의 멤버가 아니라고 추정되는 멤버의 제거는 <그림 4-12>와 같이 중위값을 기준으로 좌·우측의 밀도를 각각 산정하고, 밀도가 낮은 영역의 끝단에서 가까운 일부 멤버들을 부분적으로 소거하는 과정을 반복하면서 대표 구간통행시간 그룹의 멤버와 중심을 확정해가는 일종의 대표 그룹 중심 탐색법을 이용하게 된다. 반복횟수가 증가할수록 밀도는 증가하게 되며, 동질성(Homogeneity)이 가장 높은 그룹이 형성된다. 따라서 중위값(x_m)은 수렴하게 되며, 이때 중위값과 평균은 오차범위내에서 동일하게 된다.



<그림 4-12> 대표 그룹 중심 탐색법

대표 그룹 중심 탐색법을 적용하기 위해서는 초기 대푯값과 분포의 최댓값(x_{max})과 최솟값(x_{min})이 필요하며, 다음과 같이 산정된다. 주어진 $x = [x_n \mid n = 1, 2, \dots, N]$, $N \geq n_{min}$ 에 대하여 x_n 을 오름차순으로 정렬한 후, 구간통행시간 표본에 대한 초기 중심값 즉, 대푯값(x_m)으로 중위값(Median)을 산정한다⁷⁾. 이때 모든 x_n 은 $x_i \leq x_j, i < j, \forall i, j \in N$ 의 조건을 만족하게 되며, 최소 구간통행시간(x_{min})은 x_1 이고 최대 구간통행시간(x_{max})은 x_N 이 된다.

대표 그룹 중심 탐색법은 반복과정을 통해 지속적으로 갱신되는 x_m 을 기준으로 좌·우측 공간에 해당하는 밀도 중 밀도가 낮은 공간의 멤버를 소거해가면서 대표 그룹의 멤버에 대한 x_m 을 수렴시킨다. 소거대상 멤버는 좌·우측 공간 중 낮은 밀도의 공간을 결정한 후 선정하게 되며, 좌·우측 공간의 밀도(ρ_{lt}, ρ_{rt})는 식(4-9)와 식(4-10)으로 각각 산정된다. 여기서, N_{lt}, N_{rt} 는 각각 x_m 을 중심으로 좌측공간거리, $d_{lt} = |x_m - x_{min}|$ 와 우측공간거리, $d_{rt} = |x_m - x_{max}|$ 에 존재하는 통행시간 표본수 즉, 멤버의 개수이며, $N = N_{lt} + N_{rt}$ 이다.

$$\rho_{lt} = \frac{N_{lt}}{|x_m - x_{min}|} \quad \text{식(4-9)}$$

$$\rho_{rt} = \frac{N_{rt}}{|x_m - x_{max}|} \quad \text{식(4-10)}$$

좌·우측 공간의 밀도(ρ_{lt}, ρ_{rt})가 산정되면, 식(4-11)과 같이 낮은 밀도(ρ_{low})와 ρ_{low} 에 해당하는 거리 공간(d_{low})을 결정한다.

$$\rho_{low} = \min\{\rho_{lt}, \rho_{rt}\} \quad \text{식(4-11)}$$

7) 중위값을 이용하는 통계기반의 이상치 제거기법들은 오름차순 정렬과정을 수반할 수 밖에 없다. 따라서 대표 그룹 중심 탐색법의 초기 연산부하(Load)는 기존의 이상치제거 방법과 동일하다.

ρ_{low} 에 존재하는 통행시간 멤버 중 일부분을 소거하기 위하여 ρ_{low} 내에 존재하는 멤버의 집합을 $x_{low} = [x_{low,n} \mid n=1, 2, \dots, N_{low}]$, $1 \leq N_{low} < N$ 로 정의한다. 그리고 x 는 오름차순 정렬되었으므로 모든 $x_{low,n}$ 은 $x_{low,i} \leq x_{low,j}$, $i < j$, $\forall i, j \in N_{low}$ 의 조건을 만족한다. x_{low} 가 선정되면, 식(4-12)의 황금비(0.6180339887)을 이용하여 삭제할 멤버를 결정하게 된다. 추가로 1개 멤버를 순차적으로 삭제할 수 있지만 대푯값의 추정이라는 목적과 연산속도를 고려하였으며, x_{low} 의 일부분을 순차적으로 삭제한다.

$$\phi = \frac{\sqrt{5}-1}{2} \quad \text{식(4-12)}$$

d_{low} 과 ϕ 를 이용하여 x_m 에서 생존 임계공간(d_{trsh})을 식(4-13)과 같이 계산한 후, d_{trsh} 의 외부에 존재하는 N_{out} 개의 멤버로 구성되는 멤버집합(x_{out} , $x_{out} \in x_{low}$)을 결정한다. 이러한 밀도 기반의 임계공간 방식은 낮은 밀도에서 급격하게 공간을 좁히게 되며, 밀도가 증가함에 따라 즉 대표 통행시간 그룹이 명확해져 갈수록 점진적으로 공간을 좁혀들어간다. 따라서 대표 통행시간 그룹과 대푯값을 안정적으로 추정할 수 있게 한다.

$$\begin{aligned} & \text{IF } \rho_{lt} < \rho_{rt} \text{ then } d_{trsh} = x_m - \phi \times d_{low} \\ & \text{Else } d_{trsh} = x_m + \phi \times d_{low} \end{aligned} \quad \text{식(4-13)}$$

선정된 x_{out} 의 삭제는 식(4-14)의 규칙을 이용하여 최댓값(x_{max}) 또는 최솟값(x_{min})을 갱신하는 과정을 통해 수행된다. 여기서 x_{trsh} 는 중위값 방향으로 d_{trsh} 에서 가장 가까운 생존 멤버이다.

$$\begin{aligned} & \text{IF } \rho_{lt} < \rho_{rt} \text{ then } x_{min} = x_{trsh} \\ & \text{Else } x_{max} = x_{trsh} \end{aligned} \quad \text{식(4-14)}$$

최댓값(x_{max})과 최솟값(x_{min})이 재 설정되면, 전체 멤버의 개수를 $N=N-N_{out}$ 로 갱신한 후, $x=[x_n \mid n=1, 2, \dots, N]$ 과 대푯값(x_m)을 갱신한다. 그리고 이상에서 설명한 대표 그룹 중심 탐색법을 반복하게 되며, 종료 조건은 식(4-15)과 같다. 여기서, N_k 은 k -번째 Iteration 수행 후 N 값이며, N_o 는 초기 N 값이다. 종료조건은 초기 N 의 약 38.197%가 생존함을 의미한다. 이는 정규분포의 $u \pm \sigma$ 영역인 68.26%의 56.96%에 해당하며, 대표 통행 시간 그룹의 선정과 중심값인 대푯값에 이전의 Iteration에서 이미 도달했음을 의미한다.

$$N_k \leq (1 - \phi) \times N_o \quad \text{식(4-15)}$$

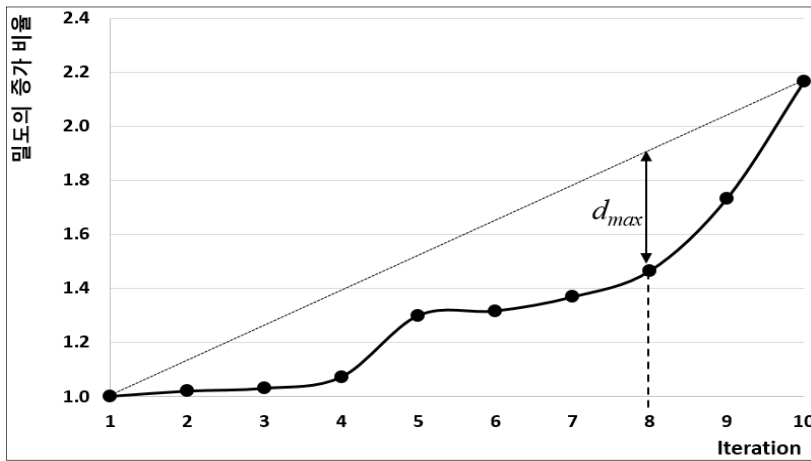
반복과정이 종료조건에 도달하면, 최적 대표 통행시간 그룹과 대표 통행시간(x_o)을 결정하게 된다. 결정은 최적 상태의 Iteration 결과를 선정하는 것이며, 다음과 같다. 대표 그룹 중심 탐색법은 낮은 밀도 상태의 멤버를 제거하면서 높은 밀도 상태를 형성한다. 따라서 Iteration이 증가함에 따라 초기 밀도에 대한 k -번째 Iteration의 밀도(ρ_k)를 이용하여 최적 상태의 결과를 도출할 수 있으며, ρ_k 는 식(4-16)과 같다. 여기서, x_{max}^o 과 x_{min}^o 는 초기 최댓값과 최솟값이고, x_{max}^k 와 x_{min}^k 는 k -번째 Iteration 수행 후 최댓값과 최솟값이며, $x_{min}^o < x_{max}^o$, $x_{min}^k < x_{max}^k$ 이다.

$$\rho_k = \frac{N_k / (x_{max}^k - x_{min}^k)}{N_o / (x_{max}^o - x_{min}^o)} = \frac{N_k \times (x_{max}^o - x_{min}^o)}{N_o \times (x_{max}^k - x_{min}^k)} \quad \text{식(4-16)}$$

<그림 4-13>은 대표 그룹 중심 탐색법의 반복연산 횟수 k 에 따른 ρ_k 의 변화를 보여주고 있다. 반복 횟수가 증가함에 따라 ρ_k 는 지수형(Exponential)으로 증가하고 있다. 따라서 k - ρ_k 관계는 전형적인 Elbow 방법(Thorndike, 1953)을 이용하여 최적의 반복회수를 결정할 수 있음을 보여주고 있으며, 최적의

반복회수는 d_{max} 일 때 이다. d_{max} 를 만족하는 $k = \{1, 2, \dots, K\}$ 는 식 (4-17)에 의해 산정되며, 여기서 K 는 식(4-15)을 만족하는 마지막 반복 회수이다. 추가로 $k-\rho_k$ 관계는 대표 그룹 중심 탐색법의 빠른 수렴과정을 설명해준다.

$$\operatorname{argmax}[(k/K \times (\max\{\rho_k\} - \min\{\rho_k\}) + \min\{\rho_k\}) - \rho_k] \quad \text{식(4-17)}$$



<그림 4-13> 대표 그룹 중심 탐색법의 반복회수-밀도변화 관계

이상에서 설명한 대표 그룹 중심 탐색법은 1번의 정렬과정을 수행하며, x_{min} , x_{max} , N , $x = [x_n \mid n=1, 2, \dots, N]$ 그리고 x_m 의 갱신에 빠른 연산을 보인다. 반면, 통계적 기법의 경우 중위값 산정을 위한 정렬과정 후, 평균과 표준편차의 산정을 위해 2번의 재 검색이 필요하다. 그리고 이상치 제거과정을 수행한 후 다시 정렬과정을 거쳐 중위값을 산정해야 한다. 따라서 대표 그룹 중심 탐색법의 수행속도는 기존의 통계기반 모형보다 빠른 연산속도를 보이며, 안정적인 통행속도 대푯값을 추정해낸다. 이상에서 설명한 대표 그룹 중심 탐색법은 <그림 4-14>의 대표 통행 시간 그룹 중심 탐색(Exploring representative group and medianoid, 이하 ERGM) 알고리즘으로 구현된다.

Given $x = [x_n \mid n = 1, 2, \dots, N]$, n_{min} at time interval $t-l$

STEP 1: Checking Minimum Sample Size

IF $N < n_{min}$ Then

-Update $(n_{min} - N)$ Samples onto x from previous time intervals

STEP 2: Initialization

- $N_o = N$

-Arrange x_n in ascending order

-Calculate initial median value(x_m^o)

-Decide initial x_{min}^o and x_{max}^o

STEP 3: Exploring the Medianoid of Potential Representative Group

Start Iteration $k = 1$ to ∞

-Calculate ρ_{lt} , ρ_{rt} by eq.(4-9) and eq.(4-10)

-Decide ρ_{low} by eq.(4-11)

-Build $x_{low} = [x_{low,n} \mid n = 1, 2, \dots, N_{low}]$, $1 \leq N_{low} < N_{k-1}$

-Calculate d_{trsh} by eq.(4-13)

-Decide x_{out} , $x_{out} \in x_{low}$ with N_{out} by using d_{trsh}

-Eliminate x_{out} from x And Then $x = x - x_{out}$

- $N_k = N_{k-1} - N_{out}$

-Update x_{min}^k and x_{max}^k by eq.(4-14)

-Calculate and Save x_m^k

-Calculate and Save ρ_k by eq.(4-16)

-Decide Stopping Condition by eq.(4-15)

IF $N_k \leq (1 - \phi) \times N_o$ Then

$K = k$

Stop Exploring

Else $k = k + 1$

STEP 4: Deciding Representative Travel Time

-Decide Optimal k_o , $k = \{1, 2, \dots, K\}$ by eq.(4-17)

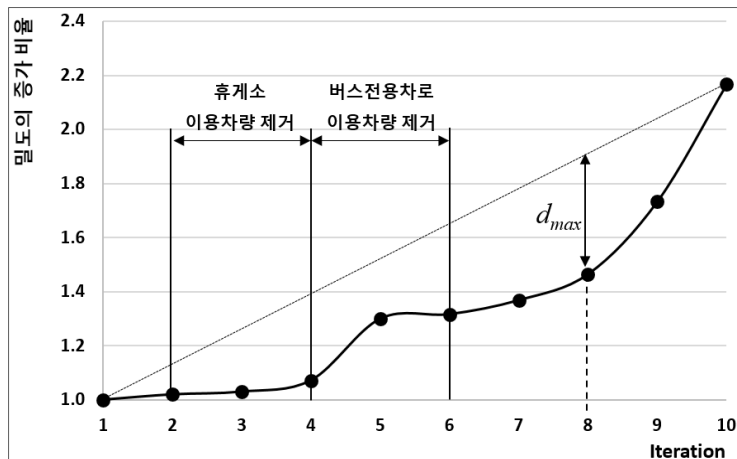
-Representative Travel Time = $x_m^{k_o}$ at time interval $t-l$

<그림 4-14> 구간통행시간 추정 알고리즘

<그림 4-15>는 ERGM 알고리즘을 이용한 대표 구간통행시간 그룹의 선정 전/후를 보여주고 있다. 버스전용차로 이용차량과 상대적으로 구간통행시간이 높은 휴게소 이용차량을 제거한 후 일반차로 이용차량 그룹을 합리적으로 선별하고 있다. <그림 4-16>은 제거 그룹에 따른 밀도의 변화를 보여주고 있다. 휴게소 이용자 그룹은 낮은 밀도에 넓게 퍼져있기 때문에 1번째 제거 그룹이 되고 있으며, 밀도가 낮기 때문에 제거과정에서 초기 밀도에 대한 k -번째 Iteration 후 밀도(ρ_k)는 급격히 증가하지 않는다. 반면, 버스전용차로 이용차량은 높은 밀도에서 그룹을 형성하고 있기 때문에 제거시 ρ_k 는 급격히 증가한다. 이후 대표 그룹의 좌·우 인접 멤버를 제거하는 과정에서 밀도는 낮게 증가하며, 대푯값인 중심점에 가까운 멤버가 제거되면서 급격하게 증가한다.



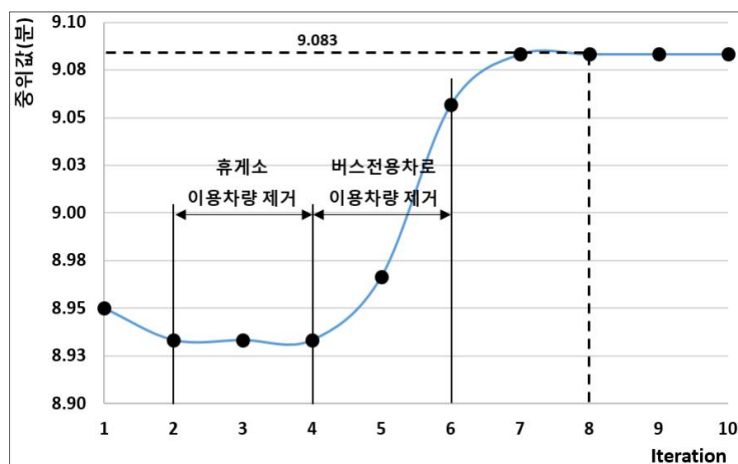
<그림 4-15> 대표 그룹 선정 전/후: 구간통행시간



<그림 4-16> 제거 대상에 따른 밀도의 변화: 구간통행시간

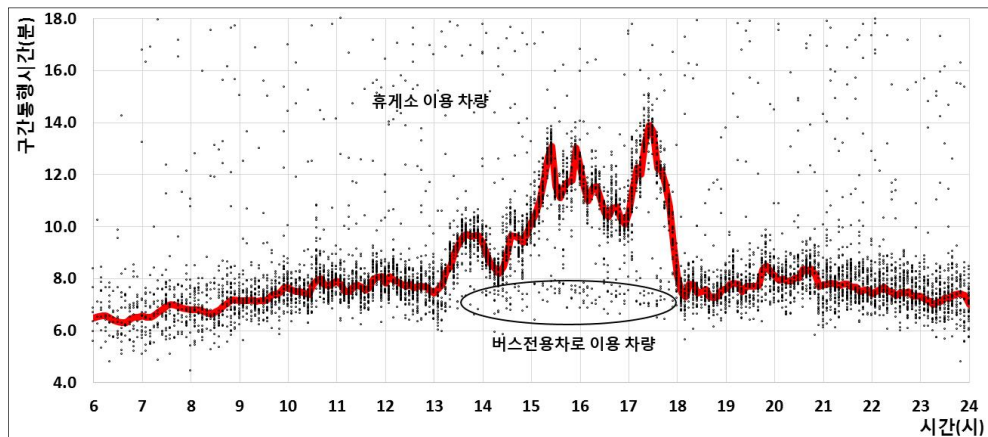
<그림 4-17>은 반복횟수에 따른 대푯값인 중위값의 변화를 보여주고 있다. 구간통행속도가 높은 휴게소 이용차량의 통행시간이 제거되면서 대푯값은 감소하며, 혼잡시 버스전용차로 이용차량의 통행시간은 일반차로 이용차량보다 낮기 때문에 낮은 대푯값을 유지한다. 이후 버스전용차로 이용차량의 통행시간이 제거되면서 대푯값은 급격히 증가한 후 수렴한다. 이후 일반차로 이용차량인 대표 구간통행속도 그룹의 멤버가 제거됨에도 불구하고 대푯값은 안정된 상태를 유지하고 있다. 이러한 사실은 ERGM 알고리즘이 구간통행시간 대푯값 추정에 있어 안정적이면서 명확한 대표 구간속도 그룹을 선별하고 있음을 의미한다.

중위값을 이용한 통계적 이상치제거 방법을 적용한 경우, 대표 통행시간(m) 8.93분, 표준편차(δ) 1.95로서 $m \pm 2\sigma$ 의 영역은 5.03~12.83으로 나타나 버스전용차로와 일반차로 이용차량을 포함한 대푯값을 산정하였다. 동질적 행태를 보이는 혼잡교통류의 구간통행시간이 $\pm 2\sigma$ 의 구간인 95.46%의 정규분포 구간에서 7.8분의 통행시간 차이는 받아들이기 어려운 사실이다. 반면, 개발 모형의 경우 대표 구간통행시간 9.08분, 표준편차 0.27로서 $\pm 2\sigma$ 의 영역은 8.54~9.62로 나타나 대표 통행시간을 기준으로 0.54분의 차이를 보이고 있으며, 중위값과 평균은 9.08로 정확히 일치하였다.



<그림 4-17> 반복회수에 따른 대푯값의 수렴: 구간통행시간

본 연구에서 통행시간 자료구축을 위해 개발된 ERGM 알고리즘은 <표 4-1>에서 설명한 휴게소 이용차량과 버스전용차로 이용차량 등의 다양한 그룹 중 대표 통행시간 그룹을 구성하는 통행시간을 합리적으로 선별하고, 안정적인 대푯값을 추정하는 것으로 판단된다. <그림 4-18>은 경부고속도로 상행선 안성휴게소 구간의 대표 구간통행시간 추정 결과를 보여주고 있으며, 통행시간의 동적 특성을 유지하면서 안정적으로 대표 구간통행시간을 추정하고 있다.



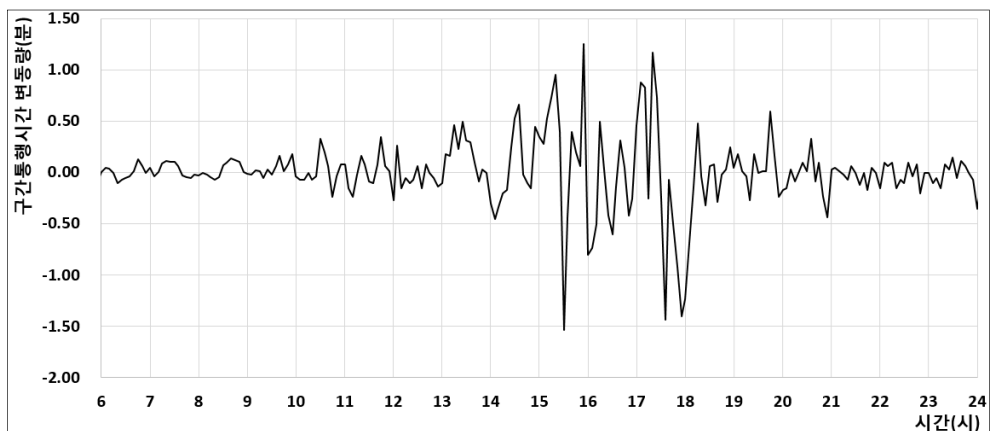
<그림 4-18> 구간통행시간 추정결과

대표 구간통행시간 추정에 있어 모형의 안정성과 더불어 주요한 요건은 통행시간의 시간적 진화상태에 대한 동적특성을 유지하는 것이다. 추정모형이 지나치게 민감한 경우 과대/과소의 추정문제를 발생시킴과 동시에 신뢰성을 저하시킨다. 반면 지나치게 둔감한 경우 상태회복⁸⁾ 시간이 길어지면서 동적특성을 상실하게된다. <그림 4-19>는 <그림 4-18>의 시간대별 변동량(=시점 $t+1$ 의 구간통행시간 - 시점 t 의 구간통행시간)을 보여

8) 본 연구에서 상태회복(State recovery)은 실제 교통상태 변화시점 이후의 상태변화를 예측 및 추정 모형이 인지하고 변화된 상태의 방향성과 변화량을 재구성함을 의미한다. 그리고 상태회복 시간은 상태회복에 소요되는 시간을 의미한다.

주고 있다. 비혼잡 시간대인 6~13시, 18~24시의 변동량은 ± 0.5 분 이내로 자연발생적 변동을 보이고 있으며 방향성의 변화시점(Turning point) 간 평균 시간길이는 2.51분, 변동량의 평균과 표준편차는 각각 -0.008, 0.177이다. 반면, 혼잡 시간대인 13~18시의 변동량은 ± 1.5 분 이내로 3배 높게, 방향성의 변화시점 간 평균 길이는 3.75분으로 1.5배 높게 나타났다. 그리고 변동량의 평균과 표준편차는 0.030, 0.583으로 분석되었다.

이상의 분석결과는 비 혼잡시의 자체변동에 비하여 혼잡시 구간통행시간 변동의 진폭과 주기는 크게 나타남을 의미한다. 이는 혼잡교통류의 동시교통류, 정체상태, 가다서다 상태가 구간통행시간에 혼재되어 있기 때문이다. 따라서 본 연구에서 개발된 구간통행시간 추정모형은 비혼잡시 구간통행시간의 자체 변동을 설명함과 동시에 안정적인 대푯값을 산출하고 있는 것으로 판단된다. 그리고 혼잡시의 안정적 상태와 불안정 상태를 설명함과 더불어 안정적이지만 동적인 특성을 보이는 대푯값을 산출하고 있다고 판단된다. 추가로 고정 상수값을 이용한 평활화기법이나 이전주기 유효통행시간 비율 등을 적용할 경우, 혼잡시 구간통행시간의 동적특성에 대한 많은 부분을 상실하게 되며, 구간통행시간 추정모형의 상태회복 시간도 증가하게 된다.

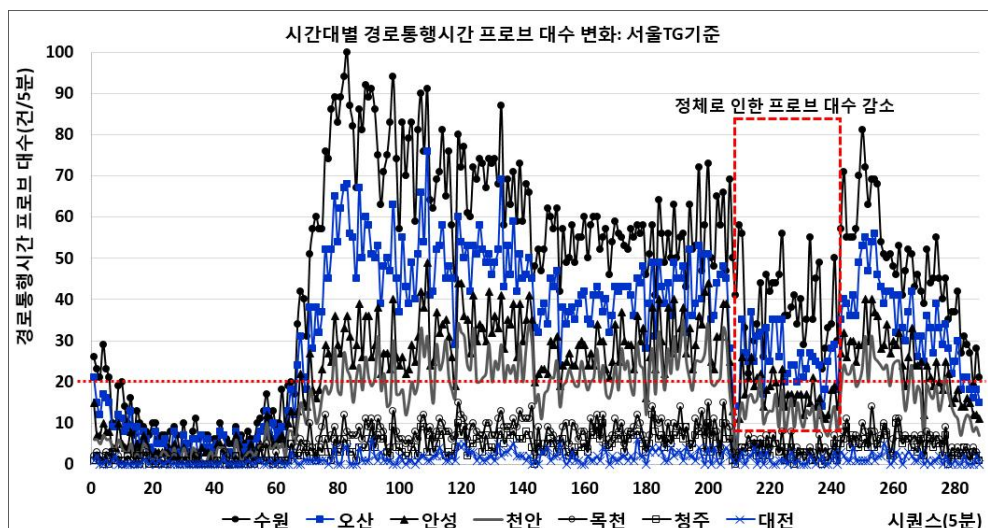


<그림 4-19> 추정 구간통행시간의 상태 변동량

3) 경로통행시간 추정 방법론 개발

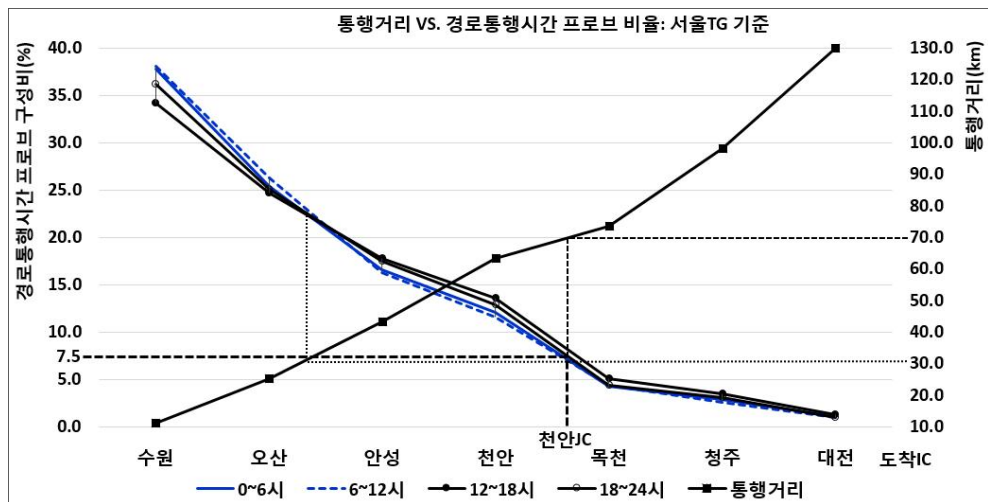
경로통행시간은 일련의 구간통행시간으로 구성되기 때문에 개별 경로통행시간 자료의 수집에 발생하는 시간차짐(Time-lag)의 길이는 구간통행시간보다 길게 나타난다. 그리고 시간차짐의 길이는 통행거리와 혼잡의 정도에 따라 증가하기 때문에 중·장거리 경로통행시간의 On-line 수집과 적용은 사실상 불가능하다. 따라서 본 연구에서 경로통행시간 추정 방법론은 과거자료 구축을 위한 Off-line 적용을 목적으로 개발하였다.

경로통행시간 프로브 표본수는 통행거리가 증가함에 따라 음지수형태로 감소하며, 이는 통행거리분포가 좌편향 분포를 가지기 때문이다. <그림 4-20>은 경부선 대전방향 서울TG를 기준의 7개 IC까지의 시간대(5분)별 경로통행시간 프로브 대수를 보여주고 있다. 프로브 대수는 1~72시퀀스(0~6시)를 제외하면 천안IC까지 대부분의 경우에 있어 20대/5분을 유지하고 있다. 그러나 천안IC이후에는 10대/5분 미만으로 급감하고 있으며, 이는 천안JC에서 통행량이 경부선과 천안-논산 고속도로로 분산되기 때문이다. 추가로 시퀀스 210~240 동안 혼잡이 심한 경우 서울TG→안성IC, 서울TG→천안IC의 프로브 대수는 20대/5분 이하로 감소하고 있다.



<그림 4-20> 통행거리 vs. 시간대별 경로통행시간 프로브 대수 변동

<그림 4-21>은 서울TG를 기준으로 통행거리에 따른 시간대별 경로통행시간 프로브 대수를 보여주고 있다. 통행거리 30km 이하의 단거리 통행 비율은 12~24시보다 0~12시가 높게, 30km 이상의 중·장거리 통행은 12~24시가 높게 나타나고 있다. 그리고 서울TG를 출발한 차량의 경로통행시간 프로브 구성비는 천안JC를 경계로 7.5%이하로 급감하고 있다. 다시말해 서울TG를 출발한 차량에 대하여 천안JC까지 7.5% 이상의 표본수를 확보할 수 있으며, 천안JC 이후로는 5% 이상의 경로통행시간 프로브를 안정적으로 확보할 수 없음을 의미한다.



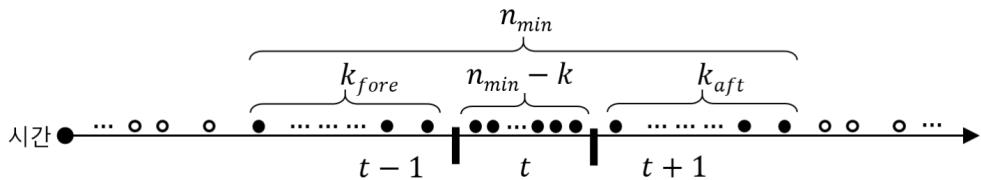
<그림 4-21> 통행거리 vs. 경로통행시간 프로브의 관계

이상의 <그림 4-20>과 <그림 4-21>을 이용하여 설명한 경로통행시간 프로브의 제약조건에서 경로통행시간을 추정하기 위하여 본 연구에서는 직접 경로통행시간 추정(Straight path travel time estimation, SPTTE)과 경유지기반 경로통행시간 추정(Relay-based path travel time, RPTTE)의 2가지 방법으로 경로통행시간을 추정하였다. SPTTE는 충분한 경로통행시간 표본수가 수집되는 단·중거리 경로에 대하여 본 연구에서 개발되어 대표 구간통행시간 추정에 적용된 ERGM 알고리즘을 동일하게 적용하는 방법이다. 그리고 RPTTE는 경로통행시간 표본수가 충

분히 수집되지 않는 장거리 경로통행시간을 추정하기 위한 방법으로서 출발지와 도착지간에 경유지를 설정하여 출발지→경유지, 경유지→출발지간 유효 표본수를 각각 확보한 후 2개의 경로통행시간을 각각 추정한 다음 추정된 경로통행시간을 하나로 결합하는 방법이며, 두 경로통행시간 추정방법에 대하여 각각 설명하면 다음과 같다.

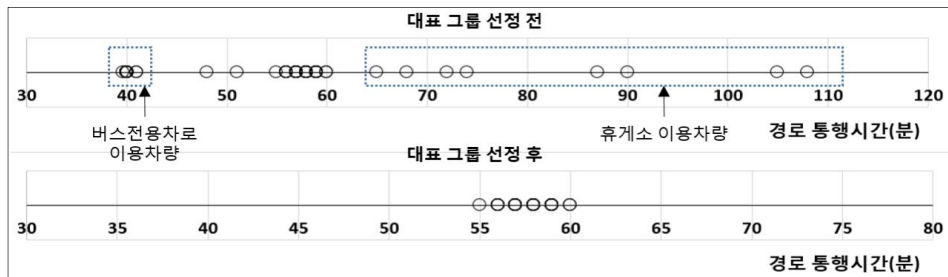
직접 경로통행시간 추정(SPTTE)은 구간통행시간 추정에 개발·적용된 대표 그룹 중심 탐색법과 동일하지만 유효 표본수 확보 방법이 다르다. 구간통행시간 추정은 구간통행시간 표본의 수집에 발생하는 시간차짐(l)을 고려한 On-line 적용을 목적으로 한다. 그러나 SPTTE는 Off-line 적용을 목적으로 하기 때문에 On-line적용시 발생하는 자료수집의 시간차짐을 고려할 필요가 없으며, 다음과 같다.

존 집합을 $z = [1, 2, \dots, Z]$ 로, $r, r \in z$ 과 $s, s \in z$ 를 각각 출발 존과 도착 존으로 정의하며, z 는 고속도로 IC의 본선과 본선에 설치된 TG로 정의한다. 그리고 시간대(t)에서 $r-s$ 간 경로통행시간($x_{r,s}$)을 산정하기 위해 수집된 개별 경로통행시간 집합을 $x = [x_n \mid n = 1, 2, \dots, N]$ 로 정의한다. $N < n_{min}$ 인 경우 <그림 4-22>와 같이 집계 시간대 t 에서 시간적으로 가까운 k 개의 경로통행시간 표본을 x 에 추가한다. 여기서 $k = k_{fore} + k_{aft}$ 이며, k_{fore} 는 t 이전(Fore)에 수집되어 추가된 표본의 개수, k_{aft} 는 t 이후에 수집되어 추가된 표본의 개수이다. 따라서 $N < n_{min}$ 인 경우 $n_{min} = N + (k_{fore} + k_{aft})$ 이며, k_{fore} 와 k_{aft} 의 개수는 시점 t 에서 k_{fore} 와 k_{aft} 의 개별 경로통행시간 출발시점간의 시간적 거리에 따라 결정된다.



<그림 4-22> 전·후방향 집계시간 확장기반 최소 표본수 확보

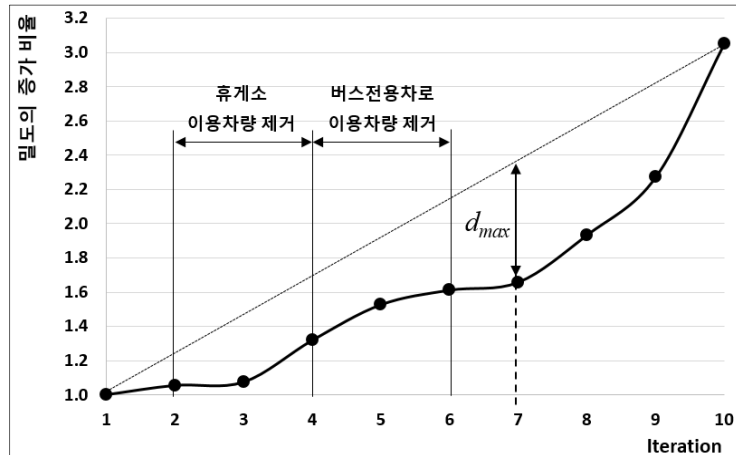
<그림 4-23>은 혼잡이 발생한 경우 ERGM 알고리즘을 이용한 대표 경로통행시간 그룹의 선정 전/후를 보여주고 있다. 경로통행시간이 높은 휴게소 이용차량과 상대적으로 경로통행시간이 낮은 버스전용차로 이용차량을 제거한 후 일반차로 이용차량 그룹을 합리적으로 추출하고 있으며, 50분 이하와 60분 이상의 타 그룹 멤버와 이상치를 제거하였다.



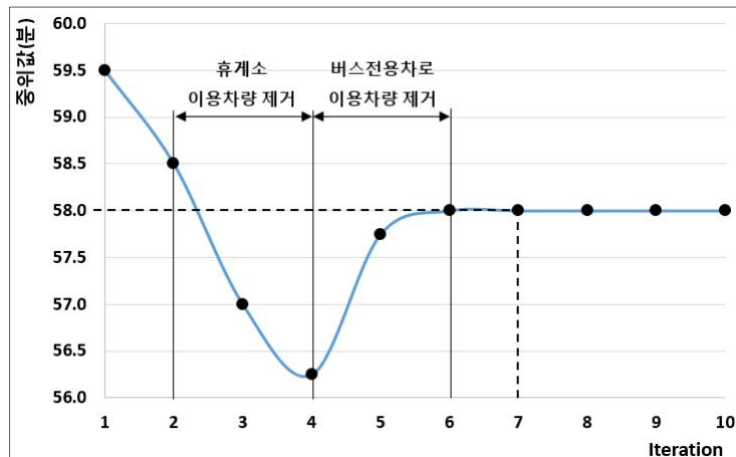
<그림 4-23> 대표 그룹 선정 전/후: 경로통행시간

<그림 4-24>는 ERGM 알고리즘의 반복횟수(k)에 따른 밀도의 변화를 보여주고 있다. 휴게소 이용차량의 경로통행시간은 일반적으로 높게 나타나기 때문에 우선 제거 대상이 되고 있으며, 밀도가 낮기 때문에 제거 과정에서 초기 밀도에 대한 k -번째 반복의 밀도(ρ_k)는 낮게 증가한다. 그러나 높은 밀도 그룹을 형성하는 버스전용차로 이용차량의 제거시 ρ_k 는 급격히 증가한다. 이후 대표 그룹의 이상치를 제거하는 과정에서 밀도는 낮게 증가한 후, 경로통행시간 대푯값인 중심점 즉, 중위값에 가까운 멤버가 제거되면서 ρ_k 는 급격하게 증가하고 있다. 그리고 식(4-17)에 의해 d_{max} 를 만족하는 최적의 반복회수(k_o)는 7로 분석되었다. <그림 4-24>는 반복회수에 따른 대푯값의 변화를 보여주고 있으며, 휴게소 이용차량과 버스전용차로 이용차량을 제거한 후 58분에서 안정적으로 수렴하고 있다. 반면, 중위값을 이용한 통계적 이상치 제거와 재 산정된 중위값을 대푯값으로 적용할 경우, 과분산이 발생하게 됨으로 버스전용차로를 이용한 일반 차량을 포함한 대표 경로통행시간을 추정하게 된다. 따라서 대표 경로통행시간은 56.25분으로 산정되며, 대표 그룹 중심 탐색법

의 추정 통행시간인 58.0분보다 2분 적은 과소추정의 문제가 발생하고 있다. 이는 통계적 방법을 적용할 경우, 경로통행시간의 절대적인 값이 구간통행시간보다 크기 때문에 과분산의 문제를 피할수 없으며 이로인한 과대/과소 추정의 문제가 발생할 수 있다는 직접적 증거이다.



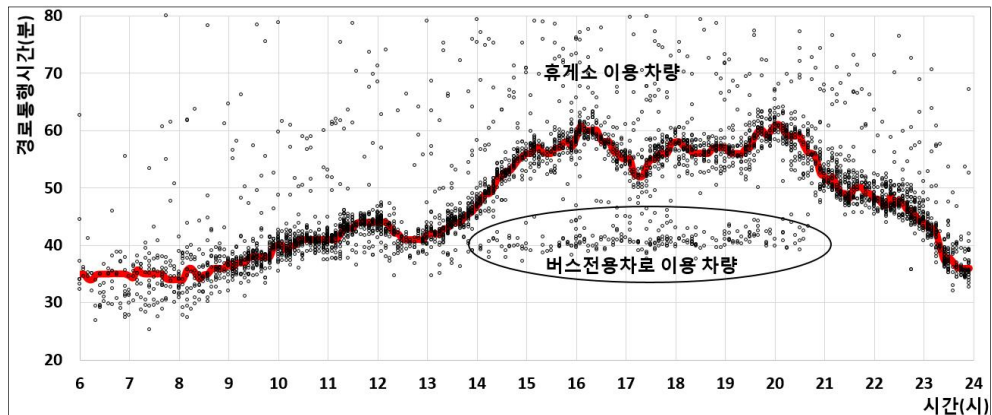
<그림 4-24> 제거 대상에 따른 밀도의 변화: 경로통행시간



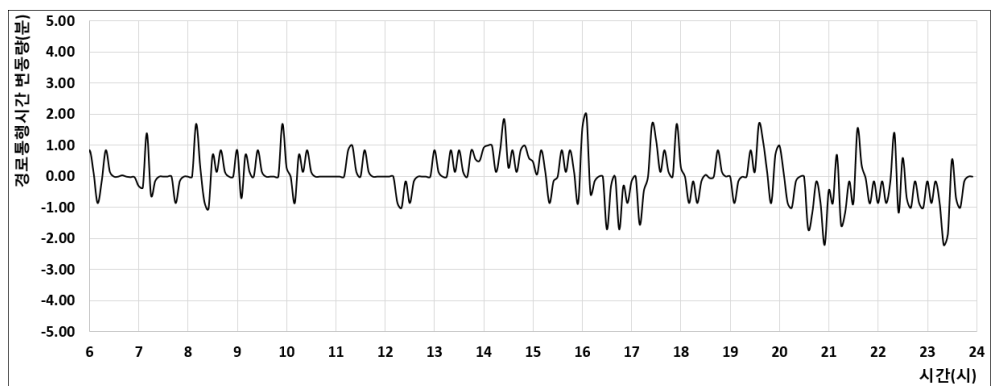
<그림 4-25> 반복회수에 따른 대푯값의 수렴: 경로통행시간

경로통행시간의 추정에 있어 주요 요건은 안정적인 대푯성과 경로통행시간의 시간적 진화상태에 대한 동적특성을 유지하는 것이다. <그림 4-26>에서 대표 그룹 중심 탐색법은 2개 그룹군, [휴게소 이용차량, 휴

게소 비 이용차량], [버스전용차로 이용차량, 휴게소 이용차량, 휴게소 비 이용을 포함한 일반차로 이용차량]에서 대표 경로통행시간 그룹을 선별 하면서 대표성을 확보하고 있다. <그림 4-27>은 <그림 4-26>의 시간대 별 상태의 변동량(=시점 $t+1$ 의 경로통행시간 - 시점 t 의 경로통행시간)을 보여 주고 있으며, 2분 이내에서 자연적인 동적특성을 보이고 있다. 참고로 집 계시간의 길이가 5분이며 자유교통류상태에서 경로통행시간이 25분이상 임으로 음의 변동량이 -5분보다 작을 경우 추정실패이다. 이는 5분 전에 출발한 차량군을 5분 후에 출발한 차량군이 5분이상 추월했다는 비 현실 적인 가정이 필요하기 때문이다.

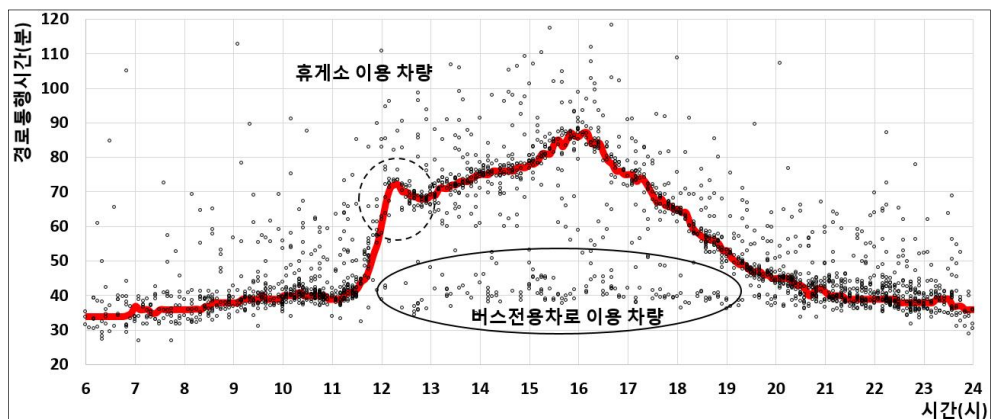


<그림 4-26> 직접 경로통행시간 추정결과: 천안IC→서울TG

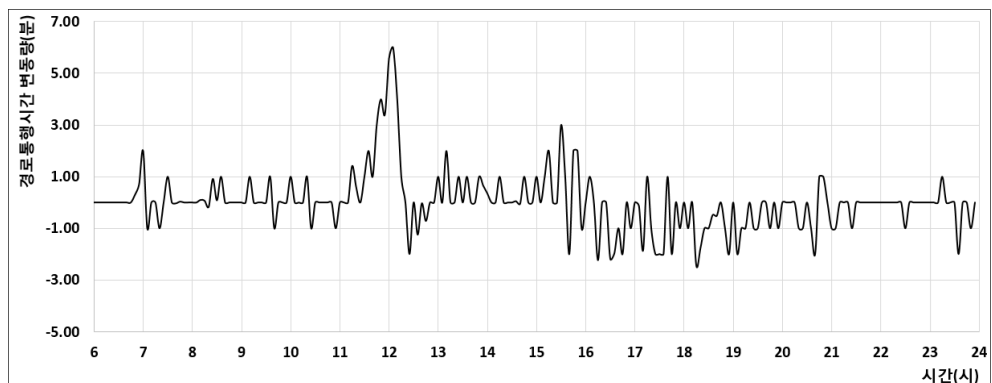


<그림 4-27> 직접 경로통행시간의 상태 변동량: 천안IC→서울TG

<그림 4-28>은 대전IC→천안IC간 SPTTE 결과를 보여주고 있다. <그림 4-26>의 천안IC→서울TG의 경우보다 낮은 프로브 대수임에도 불구하고 2개 그룹군이 명확히 구분되고 있으며, 대표 그룹 중심 탐색법은 최대 밀도에서 대표성을 확보하고 있다. <그림 4-29>의 시간대별 상태의 변동량(=시점 $t+1$ 의 경로통행시간 - 시점 t 의 경로통행시간)은 12시 전·후를 제외하면 2분 이내에서 자연적인 동적특성을 보이고 있다. 12시 전·후의 급격한 +변동량은 교통사고로 인한 높은 밀도의 국부적 혼잡 때문이며, 정상적으로 상태의 변동을 설명하고 있다. 추가로 교통사고등으로 경로통행시간 표본수가 급감하는 경우, 일반적인 통계적 기법을 적용하면 휴게소 이용차량의 영향으로 과대추정을 피하기 어렵다.

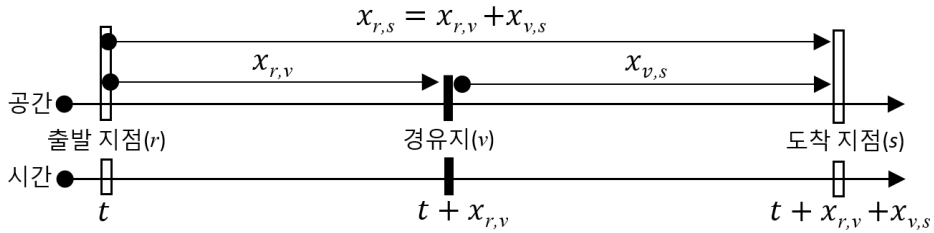


<그림 4-28> 직접 경로통행시간 추정결과: 대전IC→천안IC



<그림 4-29> 직접 경로통행시간의 상태 변동량: 대전IC→천안IC

경유지 기반의 릴레이 경로통행시간 추정(RPTTE) 방법은 유효 표본수를 확보하기 어려운 장거리 통행에 대하여 출발지(r)와 도착지(s)간에 경유지를 설정하고, 출발지→경유지, 경유지→도착지간 대표 경로통행시간을 대표 그룹 중심 탐색법을 이용하여 각각 추정한다. 그리고 추정된 2개의 경로통행시간을 시간적으로 릴레이함으로써 출발지→도착지간 경로통행시간을 산정하는 방법이다. <그림 4-30>은 경유지(v) 기반 경로통행시간 산출의 개념을 보여주고 있다. 시간대 t 에서 출발한 차량의 $r \rightarrow s$ 간 경로통행시간은 시각 t 에서 출발한 차량의 $r \rightarrow v$ 간 경로통행시간 $x_{r,v}$ 와 시각 $t + x_{r,v}$ 에서 출발한 차량의 $v \rightarrow s$ 간 경로통행시간 $x_{v,s}$ 의 합으로 산정된다. 그리고 $x_{r,v}$ 와 $x_{v,s}$ 는 SPTTE 방법론을 이용하여 추정된다.



<그림 4-30> 경유지기반 경로통행시간 산출

경유지 기반으로 장거리 $r-s$ 간 경로통행시간($x_{r,s}$)을 추정하기 위한 방법은 경유지가 존재해야 한다. 따라서 주어진 존 집합 $z = [1, 2, \dots, Z]$ 에서 연속된 여러 개의 존을 경유지 집합으로 $v = [1, 2, \dots, V]$, $v \in z$ 로 정의한다. 경유지 집합은 공간적으로 연속적인 존으로서 출발지와 도착지의 경로상에 정확히 위치한다. 여러 개의 경유지를 이용하는 이유는 구간검지체계에서 단일 존을 경유지로 설정할 경우, RSE의 장애 등으로 인한 경우 존의 경로통행시간이 수집되지 않는 경우가 발생하기 때문이다. 따라서 여러 개의 경유지를 적용함으로써 경로통행시간 수집 및 추정의 확률을 높이게 된다. RPTTE 방법의 적용은 식(4-18)의 조건을 만족할 때 적용된다. 여기서 $d_{r,s}$, $r, s \in z$ 는 $r-s$ 간 통행거리, d_{es} 는 유효 표본수의 수집이 가능한 최대거리, $z_{r,s}$, $z_{r,s} \in z$ 는 $r-s$ 간에 존재하는 z 의

집합이다. 그리고 <그림 4-20>과 <그림 4-21>의 분석결과를 기초로 d_{es} 는 70km로 설정하였다.

$$d_{r,s} > d_{es} \text{ and } v \in z_{r,s} \quad \text{식(4-18)}$$

식(4-18)을 만족하면, 경유지 집합(v)에서 경유지 j 를 선정하며, 경유지 j 는 식(4-19)의 조건에 의해 선정된다. 여기서, $d(r,j)$ 는 r 에서 j 간 거리(km) $j \in v$, $d(j,s)$ 는 j 와 s 간의 거리(km) $j \in v$, FFS_{\max} 는 자유교통류 상태에서 상위 85% 속도(kph), $x_{r,j}(t)$ 는 시각 t 에 r 에서 출발한 차량의 j 까지의 경로통행시간, $x_{j,s}(t+x_{r,j}(t))$ 는 시각 $t+x_{r,j}(t)$ 에 j 에서 출발한 차량의 s 까지의 경로통행시간이다.

$$\left\{ \begin{array}{l} \arg \min \{d(r,j)\} \\ x_{r,j}(t) > d(r,j)/FFS_{\max} \text{ and } x_{j,s}(t+x_{r,j}(t)) > d(j,s)/FFS_{\max} \end{array} \right\} \quad \text{식(4-19)}$$

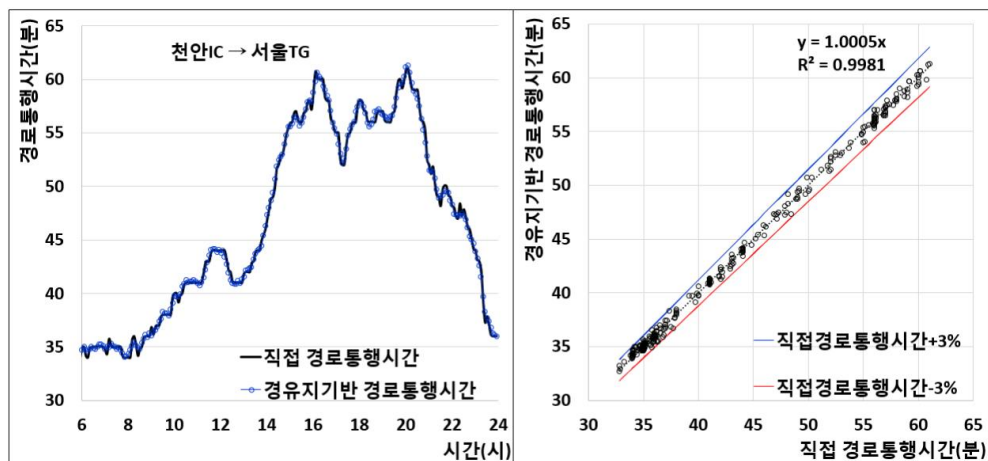
식(4-19)을 만족하는 경유지 j 가 선정되면, 식(4-20)을 이용하여 시점 t 에서 $r-s$ 간 경유지기반 경로통행시간($x_{r,s}(t)$)을 산정한다.

$$x_{r,s}(t) = x_{r,j}(t) + x_{j,s}(t+x_{r,j}(t)) \quad \text{식(4-20)}$$

RPTTE 기법은 경유지 집합을 추가하면서 통행시간 릴레이를 확장하는 경우 경로통행시간 프로브가 매우 작은 장거리 경로통행시간의 추정 가능성이 가능하다. 그러나 본 연구에서는 적용 대상 고속도로 구간인 경부고속도로 서울-대전 구간으로 한정한다. 추가로 본 연구에서 경유지 집합은 <그림 4-20>과 <그림 4-21>의 분석결과를 고려하여 천안IC와 안성IC로 설정하였으며, 천안JC에서 경로통행시간 프로브 대수가 급감/급증하기 때문에 적정한 위치선정으로 판단된다.

RPTTE 기법은 경로통행시간 프로브의 표본수가 매우 낮은 장거리 통행시간을 추정할 수 있는 장점이 있지만, RPTTE의 추정력을 평가하기 위한 참값에 가까운 장거리 통행시간 대푯값의 수집은 용이하지 않다.

따라서 실제 장거리 통행시간과 RPTTE 통행시간을 비교·검증하기 어려운 실정이다. 따라서 프로브 수집율이 높은 천안IC→서울TG 구간을 장거리 통행으로 가정하고 오산IC를 경유지로 설정한 후 릴레이 경로통행시간(RPTT) 추정결과를 검증하였으며, <그림 4-31>과 같다. 직접경로통행시간(SPTT)과 RPTT는 시계열적으로 매우 유사한 행태를 보이고 있으며, SPTT의 국부적 변동을 유지하고 있다. RPTT의 SPTT $\pm 3\%$ 이내 적중률은 100%로 나타났으며, 절대평균오차는 0.306분으로 분석되었다. 쌍체 T-검정결과 ($\alpha=0.05$) t값은 -1.189로서 기각역인 1.960보다 작으며, 유의확률은 0.235로서 유의수준 0.05보다 매우 큼으로 귀무가설을 기각할 수 없다. 따라서 RPTT와 SPTT는 통계적으로 다르지 않다.



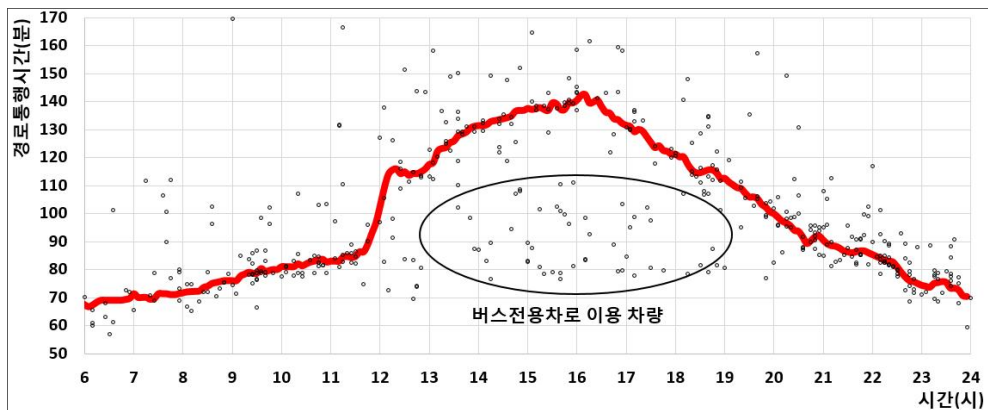
<그림 4-31> 직접 경로통행시간과 경유지 기반 경로통행시간의 비교

대응표본 검정									
		대응차				t	자유도	유의확률 (양쪽)	
		평균	표준편차	평균의 표준오차	차이의 95% 신뢰구간				
					하한				상한
대응 1	SPTTE - RPTTE	-.02746	.39187	.02309	-.07291	.01799	-1.189	287	.235

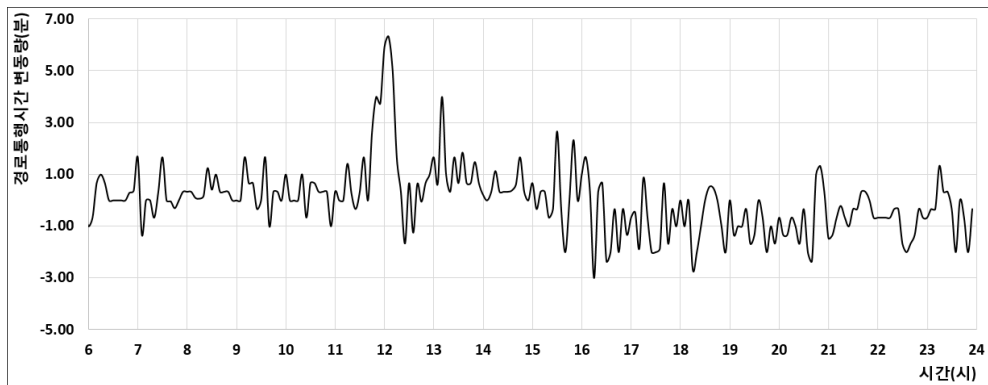
<그림 4-32> 직접 경로통행시간과 경유지 기반 경로통행시간의 검정 결과

<그림 4-33> RPTTE 기법으로 추정된 대전IC→서울TG의 경로통행시간을 보여주고 있으며, 경유지는 천안IC이다. 버스전용차로 이용차량을 제외하더라도 대표 통행시간의 중심점은 명확히 구분되고 있지 않다. 따

라서 일반적인 통계기반의 추정기법을 적용할 경우 심각한 과대/과소 추정
 정의 문제가 발생한다. 엄밀히 말해, 현재 적용되고 있는 통행시간 추정
 모형을 적용할 수 없다고 판단되며, 이는 분포내에서 고밀도의 중심점을
 보이지 않기 때문이다. 반면, RPTTE로 추정된 경로통행시간은 <그림
 4-26>과 <그림 4-28>을 고려하여 직관적으로 판단할 때 합리적인 대푯
 값이라고 판단된다. 이전 시간대에 대한 상태변동량은 12시 전·후를 제
 외하면 ± 3 분이내에서 자체 변동을 보이고 있으며, 12시 전·후의 급격한
 +변동은 대전IC→천안IC구간의 교통사고 영향이다. 따라서 RPTTE 기법
 은 저밀도 상태에서 효과적으로 대표 경로통행시간을 추정할 수 있는 방
 안중 하나라고 판단된다.



<그림 4-33> 경유지기반 경로통행시간 추정결과: 대전IC→서울TG



<그림 4-34> 경유지기반 경로통행시간의 상태 변동량: 대전IC→서울TG

이상에서 설명한 직접 경로통행시간 추정(SPTTE) 방법과 경유지기반 경로통행시간 추정(RPTTE) 방법은 <그림 4-35>의 경로통행시간 추정 알고리즘으로 구현된다. 알고리즘은 <그림 4-14>의 ERGM 알고리즘을 적용하는 SPTTE 단계와 SPTTE 단계에서 추정된 직접 경로통행시간 (SPTT)를 이용하여 경유지기반 릴레이 경로통행시간(RPTT)을 추정하는 RPTTE 단계로 구성된다.

Given $z = [1, 2, \dots, Z]$, $r, s \in z$, $d_{r,s}$, d_{es}

SPTTE STEP: Estimating Short-and-Middle-Distance Path Travel Time
For all $r-s$ paths and all time intervals

IF $d_{r,s} \leq d_{es}$ Then

STEP 1: Checking Minimum Sample Size
Given $x = [x_n \mid n = 1, 2, \dots, N]$, n_{min} at time interval t
IF $N < n_{min}$ Then
-Update $(n_{min} - N)$ Samples onto x from fore-and-aft time intervals

STEP 2: Estimating Path Travel Time Using ERGM Algorithm
-Initialization using ERGM STEP 2
-Exploring the Medianoid of Potential Representative Group using ERGM STEP 3
-Deciding Representative Travel Time using ERGM STEP 4
-Save $r-s$ SPTT at time interval t

RPTTE STEP: Estimating Long-distance Path Travel Time
Given $v = [1, 2, \dots, V]$ $v \in z$, $z_{r,s}$ $z_{r,s} \in z$,
 $r-s$ SPTT $\forall r, s, \forall \text{time intervals}$

For all $r-s$ paths and all time intervals

IF $d_{r,s} > d_{es}$ and $v \in z_{r,s}$ given eq.(4-18) Then

-Decide passage(j), $j \in v$ using eq.(4-19)

IF $\left\{ \begin{array}{l} \arg \min \{d(r, j)\} \\ x_{r,j}(t) > d(r, j)/FFS_{\max} \text{ and } x_{j,s}(t + x_{r,j}(t)) > d(j, s)/FFS_{\max} \end{array} \right. = \text{TRUE}$ Then

-Estimate $x_{r,s}(t)$ by eq.(4-20)

<그림 4-35> 경로통행시간 추정 알고리즘

5. 자료 구축

1) 수집자료 선정

본 연구에서 개발한 KJC 예측기는 KNN-NPR 모형을 근간으로 하며, 자료수집에 시간차집이 발생하는 구간검지체계하에서 수집되는 자료에 중점을 두고 개발되었다.⁹⁾ 그리고 KJC 예측기는 도로망의 소통상태를 예측한 후, 예측된 소통상태를 이용하여 단·중·장거리 경로통행시간을 예측하게 된다. 추가로 KNN-NPR모형은 이력 데이터베이스에 내재된 정보에 전적으로 의존한다(Davis and Nihan, 1991; Charytoniuk et al., 1998). 즉, 데이터에 내재된 다양성을 확보하기 위해서는 데이터의 양과 더불어 질이 중요하다. 따라서 개발모형의 성능평가에 이용되는 자료는 단·중·장거리 통행을 모두 포함하는 도로망의 구간검지체계를 통해 수집되어야 하며, 수집되는 자료의 양과 질을 담보할 수 있어야 한다. 그리고 개발모형의 평가에 대한 도로망의 대표성을 포함해야 한다. 추가로 소량의 지점 또는 단일 구간의 자료를 이용한 KNN관련 일부의 연구는 이력자료에 내재된 다양성을 확보하지 못 하였기 때문에 예측모형의 성능은 저하되었다(Chang et al. 2012, Yoon and Chang, 2014).

이상의 고려를 통해 자료의 신뢰성이 높은 한국도로공사의 DSRC시스템과 TCS시스템에서 수집된 자료를 이용하여 데이터의 양과 질을 담보함과 더불어 다양성을 확보하였다. 그리고 중·장거리 통행은 지역간 도로에서 높게 나타난다. 일반적으로 지역간 고속도로는 도시부 반복적 통행특성과 달리 다양한 통행특성을 보이며, 이는 다양성을 간접적으로 설명해 준다고 하겠다. 따라서 개발모형의 평가에 대한 도로망의 대표성과 중·장거리 통행을 고려하여 대상 고속도로는 경부고속도로 서울-대전 구간으로 선정하였다.

9) 본 연구에서 개발된 KJC예측기는 구간검지체계와 지점검지체계에서 수집되는 교통 시계열자료에 적합하도록 설계되었으나, 일반적인 도로망 적용이라는 관점에서 구간검지체계에 중점을 두었다.

2) 수집자료의 내용 및 범위

본 연구에서 개발된 KJC 예측기의 최종 목표는 도시부와 지방부의 도로망 소통상태를 예측하고 이를 이용하여 중·장거리 경로통행시간을 예측하는데 있다. 따라서 공간적 범위는 단·중·장거리 통행이 모두 포함되는 경부고속도로 대전IC-양재IC구간(142.3km)를 선정하였다. 해당구간에는 JC 8개소, IC/TG 12개소, 고속도로 휴게소 13개소가 운영되고 있다. 이는 고속도로 통행행태의 다양성이 수집자료에 내재되어 있음을 간접적으로 의미한다. 해당 고속도로 구간은 통행행태의 측면에서 도시부 고속도로의 특성과 지역간 고속도로의 특성을 모두 포함하고 있다. 그리고 교통운영의 측면에서 버스전용차로, 가변차로시스템 등 다양한 운영기술이 적용되고 있는 구간이다. 따라서 앞장에서 언급한 이력자료의 다양성을 확보할 수 있을 것으로 판단된다.

시간적 범위는 2010년 8월 1일 ~ 2011년 03월 31일까지 총 245일이다. 수집대상 자료는 구간통행시간, 구간통행속도, 그리고 구간 프로브통행량 자료구축을 위한 개별 DSRC 프로브 자료, 그리고 도로망 유·출입 통행량 자료구축을 위한 TCS 유·출입 통행량 자료이다. 추가로 구간자료 구축을 위한 구간의 경계 즉, 노드를 구성하기 위하여 RSE(노변기지국) 이점정보를 수집하였다. DSRC자료는 대상 고속도로 구간에 존재하는 RSE를 경유한 차량의 개별 통행사슬(Trip chain) 자료를 수집하였으며, 수집된 자료의 크기는 1일 평균 157만건으로서 전체 자료의 크기는 약 3억 8천 500만건이며, 약 48GB이다. 그리고 TCS자료는 총 12개 TG에 대한 유·출입 방향별 자료를 수집하였다.

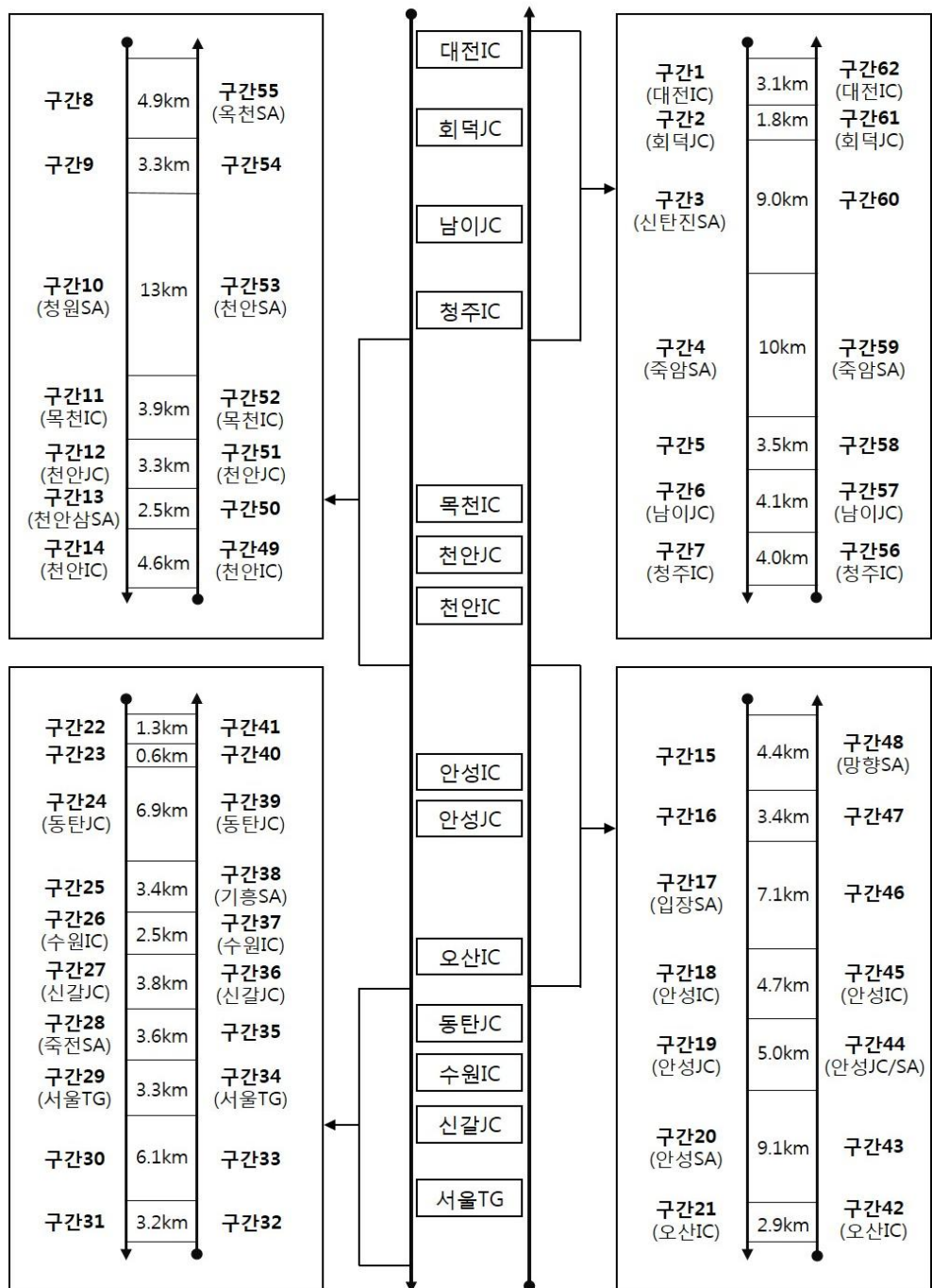
3) 자료구축을 위한 설계

자료를 시·공간적으로 집계·구축하기 위해서는 도로구간 및 존, 대상차종, 그리고 집계시간대 길이가 설정되어야 한다. 도로구간 및 존은 <표 4-2>와 같이 양방향 62개 구간과 8개 존으로 설정하였으며, 도로구간별 시설물 세부정보는 <그림 4-36>과 같다. 설정된 도로구간의 길이

는 0.6~13km이며, 평균길이는 4.59km이다. 선정된 인접 존간 길이는 10.23~32.81km이며, 평균길이는 18.71km이다.

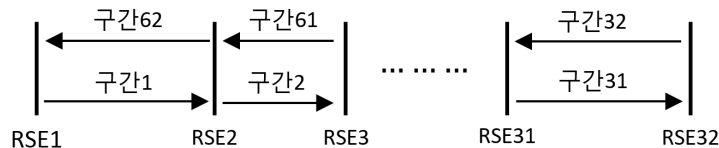
<표 4-2> 도로구간 및 존 설정

상행 (서울방향)		하행 (부산방향)		존(이정)	거리 (km)
구간	시설	구간	시설		
1		62		대전IC (272. 30)	3. 1
2	회덕JC	61	회덕JC		1. 8
3	신탄진SA	60			9. 0
4	죽암SA	59	죽암SA		10. 0
5		58			3. 5
6	남이JC	57	남이JC		4. 1
7		56		청주IC (305. 11)	4. 0
8		55	옥천SA		4. 9
9		54			3. 3
10	청원SA	53	천안SA		13. 0
11		52		목천IC (329. 91)	3. 9
12	천안JC	51	천안JC		3. 3
13	천안삼SA	50			2. 5
14		49		천안HC (340. 14)	4. 6
15		48	망향SA		4. 4
16		47			3. 4
17	입장SA	46			7. 1
18		45		안성IC (360. 11)	4. 7
19	안성JC	44	안성JC/SA		5. 0
20	안성SA	43			9. 1
21		42		오산HC (378. 21)	2. 9
22		41			1. 3
23		40			0. 6
24	동탄JC	39	동탄JC		6. 9
25		38	기흥SA		3. 4
26		37		수원IC (392. 78)	2. 5
27	신갈JC	36	신갈JC		3. 8
28	죽전SA	35			3. 6
29		34		서울TG (403. 30)	3. 3
30		33			6. 1
31		32			3. 2



<그림 4-36> 설정된 도로구간의 세부정보

도로구간은 RSE의 위치를 기준으로 <그림 4-37>과 같이 구분하였다. 그리고 대상 도로구간에 운영중인 총 40개의 RSE중 자료수집기간에 운영상태가 양호한 32개의 RSE를 선정하였다.



<그림 4-37> 도로구간 설정

경로통행시간 자료구축을 위한 존은 주요 IC와 TG를 대상으로 8개 존을 선정하였으며, 존간 거리는 <표 4-3>과 같다. 그리고 존의 중심은 IC/TG가 위치한 고속도로의 본선으로 설정하였다.

<표 4-3> 선정된 존간 거리(km)

존명	대전IC	청주IC	목천IC	천안IC	안성IC	오산IC	수원IC	서울TG
대전IC	-	32.7	57.5	67.7	87.7	105.8	119.9	130.9
청주IC	32.7	-	24.8	35.0	55.0	73.1	87.2	98.2
목천IC	57.5	24.8	-	10.2	30.2	48.3	62.4	63.2
천안IC	67.7	35.0	10.2	-	20.0	38.1	52.2	63.2
안성IC	87.7	55.0	30.2	20.0	-	18.1	32.2	43.2
오산IC	105.8	73.1	48.3	38.1	18.1	-	32.2	25.1
수원IC	119.9	87.2	62.4	52.2	32.2	14.1	-	11.0
서울TG	130.9	98.2	73.4	63.2	43.2	25.1	11.0	-

통행시간과 통행속도 자료의 구축을 위해 선정된 대상 차종은 한국도로공사의 요금징수 6개 차종 중 승용차(1종)과 경차(6종)이다. 그리고 도로망 유·출입 통행량 자료구축을 위한 TCS 유·출입 통행량 자료는 모든 차종을 대상으로 하였다. 일반적으로 교통정보제공을 위한 장래상태의 예측은 집계시간길이의 n 배로 정해지며, 보통은 교통정보의 신속한

제공을 위하여 동일하게 적용하는 경우가 대부분이다. 기존연구의 고찰 결과, 교통자료의 동적 특성을 유지하기 위한 집계 시간길이는 20초~15분까지 다양하게 제시되었다. 교통자료의 수집주기에 대한 몇몇 연구가 보고되었지만, 적정 교통정보 제공주기에 대한 명확한 이론적 근거는 제시되고 있지 않은 실정이다. 그리고 국내 ATIS관련 시스템의 교통정보 제공주기는 통상 5분을 적용하고 있다. 이상의 이론과 실제 시스템의 운영상태를 고려하여 자료의 집계시간 길이와 예측모형의 시간대 길이(TL)는 교통정보 제공주기를 고려하여 5분으로 설정하였다.

4) 자료구축의 과정

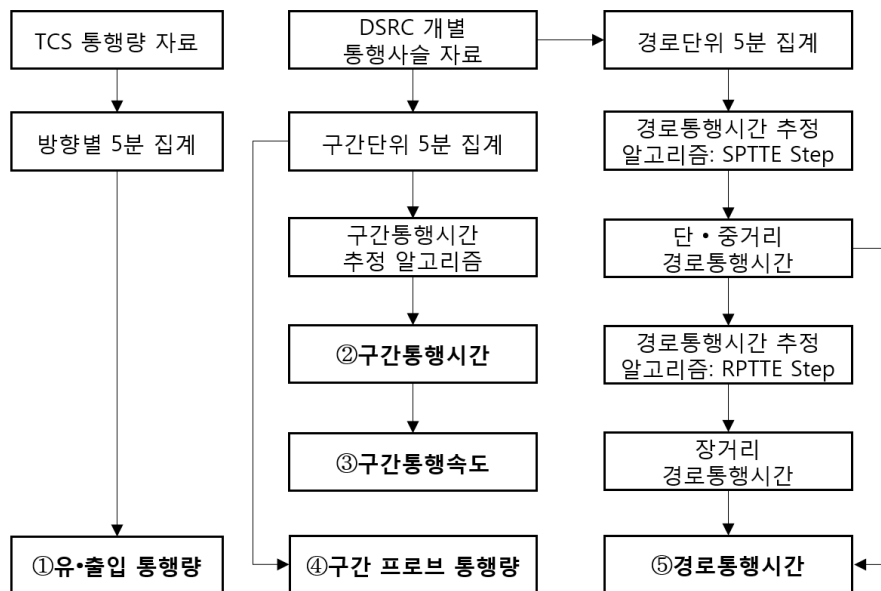
자료구축은 도로구간 및 존, 대상차종, 그리고 집계시간대 길이를 기준으로 지점단위, 구간단위 그리고 경로단위 자료로 구축하였으며, <그림 4-38>과 같다. TCS의 모든 차종을 대상으로 선정된 12개 영업소별 유입/유출 방향별 일별 시퀀스¹⁰⁾별 도로망 유·출입 통행량 자료를 구축하였다. 따라서 구축된 유·출입 통행량의 자료크기는 $1,693,440(=12\text{개소} \times 2\text{개방향} \times 245\text{일} \times 288\text{시퀀스/일})$ 건이다.

구간단위 자료는 구간통행시간, 구간통행속도, 그리고 구간 프로브 통행량을 각각 구축하였다. 통행시간과 통행속도는 승용차(1종)과 경차(6종)를 대상으로 하였으며, 구간 프로브 통행량은 모든 차종을 대상으로 하였다. 구간단위 자료의 구축은 RSE의 ID와 통과시각을 기준으로 DSRC 개별 통행사슬(Trip chain) 자료를 이용하여 단계별로 구축하였다. 구간 프로브 통행량 자료는 5분집계로 62개 구간에 대하여 일별로 구축하였으며, 구축된 자료의 크기는 $4,374,720(=62\text{개구간} \times 245\text{일} \times 288\text{시퀀스/일})$ 건이다. 구간통행시간 자료는 대상차종인 승용차와 경차의 개별 구간통행시간을 이용하여 최소 표본수를 만족하는 5분 집계자료를 구축한 후, 본 연구에서 개발된 <그림 4-14>의 구간통행시간 추정 알고리

10) 집계 시간대 길이를 5분으로 설정하였으므로, 데이터베이스의 시간대 개수인 시퀀스는 1일에 $288(=60\text{분}/5\text{분} \times 24\text{시간})$ 개이다.

즘을 이용하여 62개 구간별 245일 288개 시퀀스에 대하여 구축하였다. 구축된 구간통행시간 자료와 구간거리를 이용하여 구간통행속도 자료를 구축하였으며, 구간통행시간과 구간통행속도 구축자료의 크기는 각각 $4,374,720(=62\text{개구간} \times 245\text{일} \times 288\text{시퀀스/일})$ 건이다.

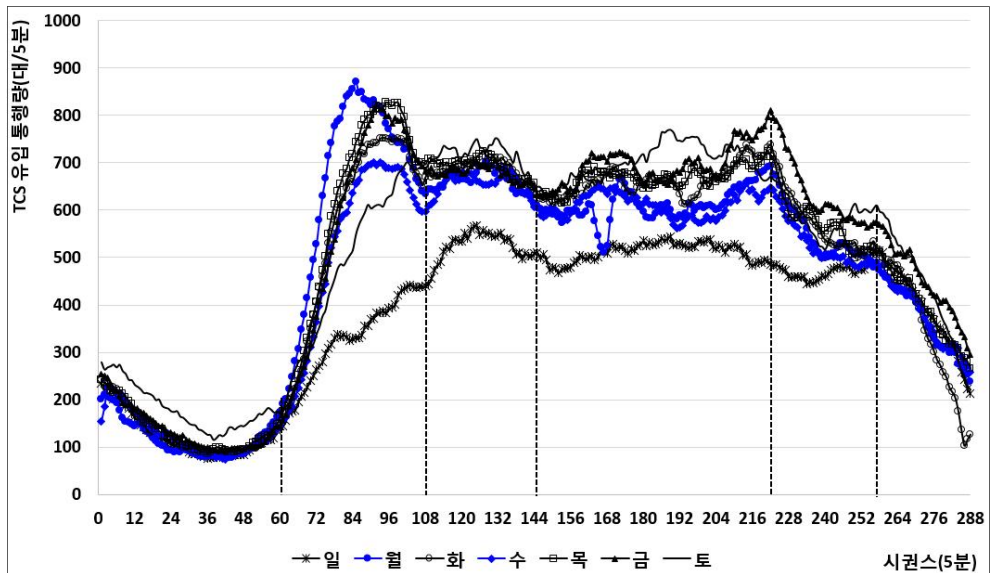
경로단위 자료는 존간 경로통행시간이며, 승용차와 경차를 대상으로 선정된 8개 존에 대한 56개 경로쌍($8 \times (8-1)$)에 대하여 일별 시간대별로 구축하였으며, 구축과정은 다음과 같다. 각 경로쌍별 5분 집계자료를 구축한 후 본 연구에서 개발된 <그림 4-35>와 같은 경로통행시간 추정 알고리즘의 SPTTE Step을 이용하여 단·중거리 경로통행시간자료를 구축하였다. 구축된 단·중거리 경로통행시간 자료와 경로통행시간 추정 알고리즘의 RPTTE Step을 이용하여 장거리 경로통행시간 자료를 구축하였다. 그리고 구축된 경로통행시간 자료의 크기는 총 $3,951,360(=56\text{개 경로쌍} \times 245\text{일} \times 288\text{시퀀스/일})$ 건이다. 따라서 구축된 5개 자료의 크기는 총 18,768,960건이다.



<그림 4-38> 세부 자료별 구축과정

5) 유·출입 통행량 구축자료의 특성

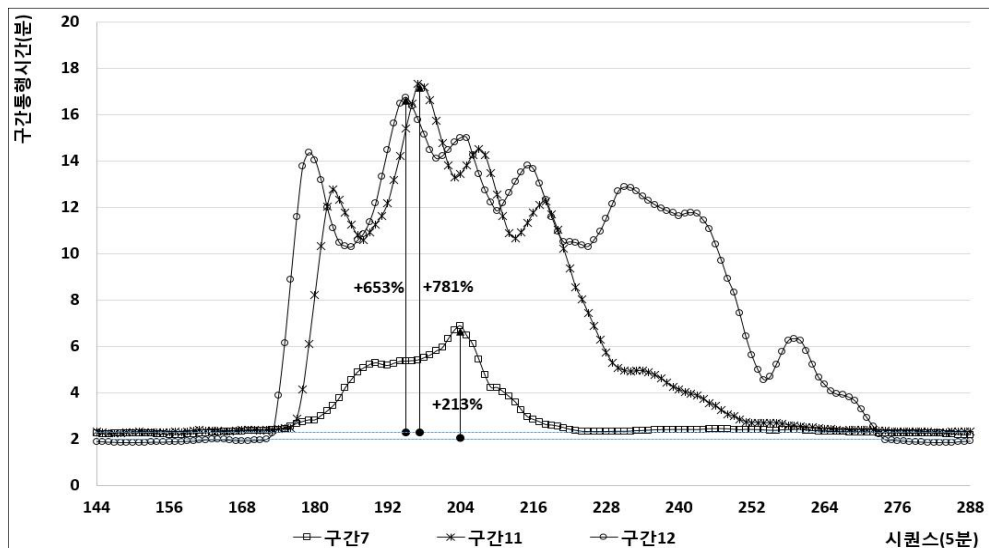
유·출입 통행량은 도로망의 반복적 혼잡을 발생시키는 주요 원인 중 하나이다. 따라서 유·출입 통행량의 요일별 시간대별 패턴은 주요한 입력변수이다 [참고 <그림 3-11>]. <그림 4-39>는 구축된 유입 통행량의 요일별 시간대별 시계열 변동을 보여주고 있다. 요일별 특성이 뚜렷히 나타나고 있으며, 3~5개의 첨두특성을 가진다. 토요일의 0~5시 일요일의 5~24시는 뚜렷이 구분되고 있다. 오전 시간대 통행량은 심야시간대에 대비 689.5~912.8%까지 증가하고 있다. 시퀀스별 변동은 108시퀀스 이전과 220시퀀스 이후를 기준으로 안정적인 변동을 보이고 있으나, 108~220 시퀀스의 경우 추세는 안정적이나 국부적 변동은 불안정하게 나타나고 있다. 시계열의 추세가 안정적이고 국부적 변동이 불안정한 경우, KNN기반 예측모형의 입력자료 시간대 개수(l)는 증가하게 되며, 모형의 정확도는 저하된다. 그러나 개발모형은 다수의 유·출입 통행량을 이용하기 때문에 국부적 영향은 크지 않을 것으로 판단된다.



<그림 4-39> TCS 통행량 구축자료의 특성

6) 구간통행시간 구축자료의 특성

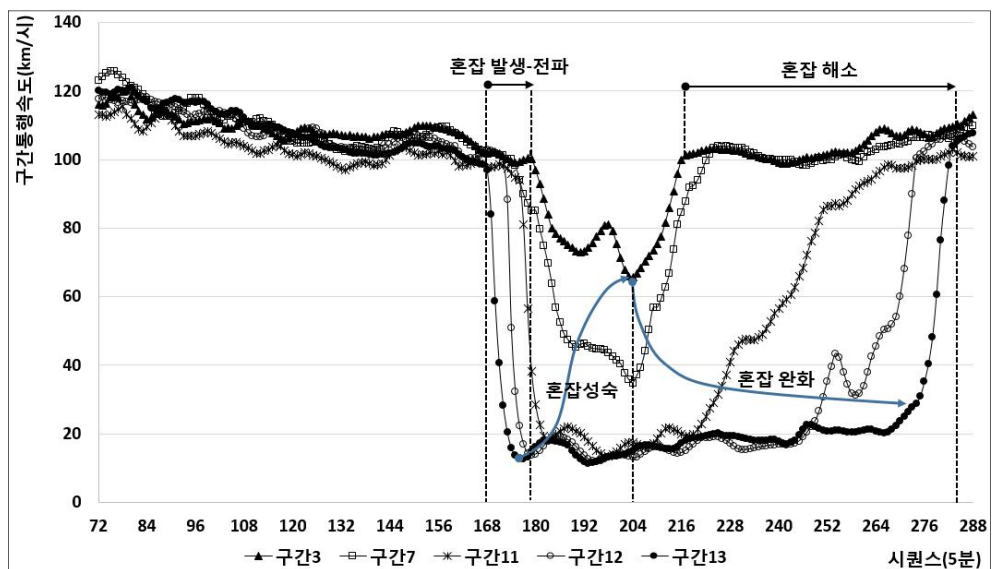
구간통행시간은 현행상태 입력자료의 수집에 발생하는 시간처짐의 개수(τ)를 결정하기 때문에 장래상태의 예측력과 관련성이 매우 높다. <그림 4-40>에서 혼잡상태의 통행시간은 정상상태에 비하여 213~781%까지 증가하고 있으며, 다음의 3가지 측면을 말해준다. 첫째, 충분한 표본수를 이용하였기 때문에 상태의 국부적 변동은 실제로 크지 않으며, 안정적 상태와 상태 전환시가 명확히 구분되고 있다. 둘째, 통행시간은 최대 7~17분까지 증가하고 있다. 따라서 자료수집에 발생하는 τ 는 5분 집계시간대로 고려할 경우 1~4까지 변화하고 있다. 실제로 이상치를 제외한 모든 프로브를 수집하기 위해서 식(3-2)를 적용할 경우 τ 는 1~5까지 증가한다. 따라서 장래시점 $t+1$ 을 예측하기 위해서는 시각 $t-\tau$ 에서 $\tau+1$ 의 예측해야 하며, 이는 기존 단기 예측모형의 예측영역을 초과한다. 마지막으로 혼잡시 구간통행시간은 비혼잡시에 비하여 213~781%까지 변화하고 있다. 이는 과거 속도 Profile 등과 같은 정적인 평균기반의 예측값을 적용할 경우 예측오차는 112.5~390.5%까지 증가할 수 있음을 의미한다. 추가로 연속류의 추정된 구간통행시간이 빈번한 방향전환을 보이는 경우 수집자료의 정확도 또는 추정모형의 문제일 가능성이 매우 높다.



<그림 4-40> 구간통행시간 구축자료의 특성

7) 구간통행속도 구축자료의 특성

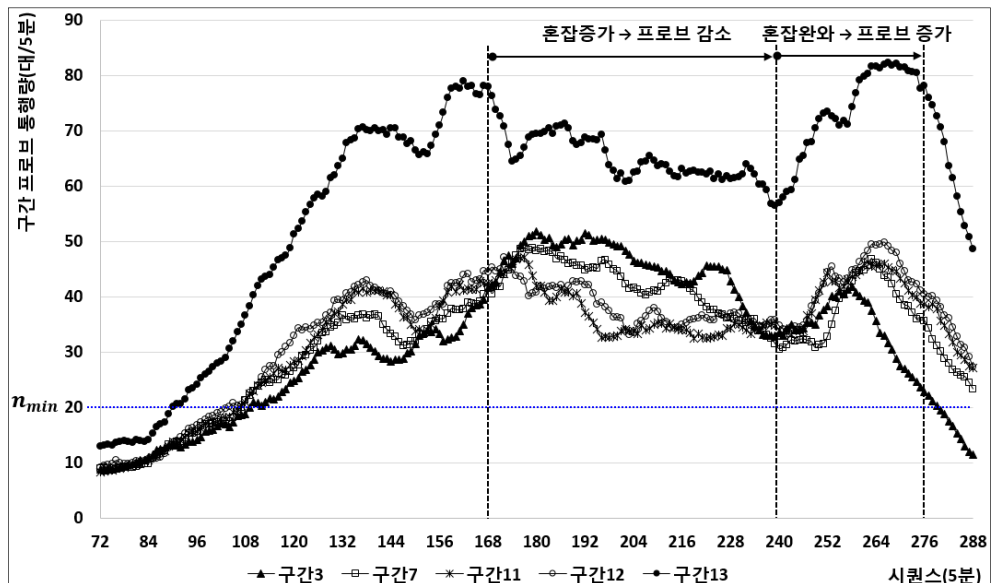
교통사고 등으로 인한 비반복적 행태를 제외하면 연속류의 구간통행속도는 <그림 4-41>과 같이 10~120km/시에서 최대/최소 속도를 보이며, 최대속도는 정상교통류상태에서 교통량이 증가함에 따라 120에서 100km/시까지 감소하고 있다[참고: 구간13에서 구간3의 거리는 59km임]. 구간13에서 발생한 혼잡의 구간3까지 전파시간은 60분으로 나타나 1차 충격파의 속도는 약 60km/시이다. 이는 하루부 구간에서 발생한 혼잡을 예측모형이 인지하는데 12개의 시간대라는 충분한 여유가 있음을 말해준다. 성숙된 혼잡의 전파시간은 140분이 소요되어 2차 고밀도 충격파의 속도는 약 25km/시로 분석되었으며, 완화된 혼잡의 공간적 진화에 총 355분이 소요되어 저밀도 전방충격파의 속도는 약 10km/시로 분석되었다. 그리고 혼잡의 완전 해소 상태 전파시간은 약 355분이 소요되어, 저밀도 전방충격파의 속도는 약 10km/시로 나타났다. 이상의 분석결과는 도로망의 소통상태를 입력자료로 적용할 경우 혼잡의 발생-성장-완화-소멸 과정을 사전에 예측할 수 있음을 의미한다.



<그림 4-41> 구간통행속도 구축자료의 특성

8) 구간 프로브 통행량 구축자료의 특성

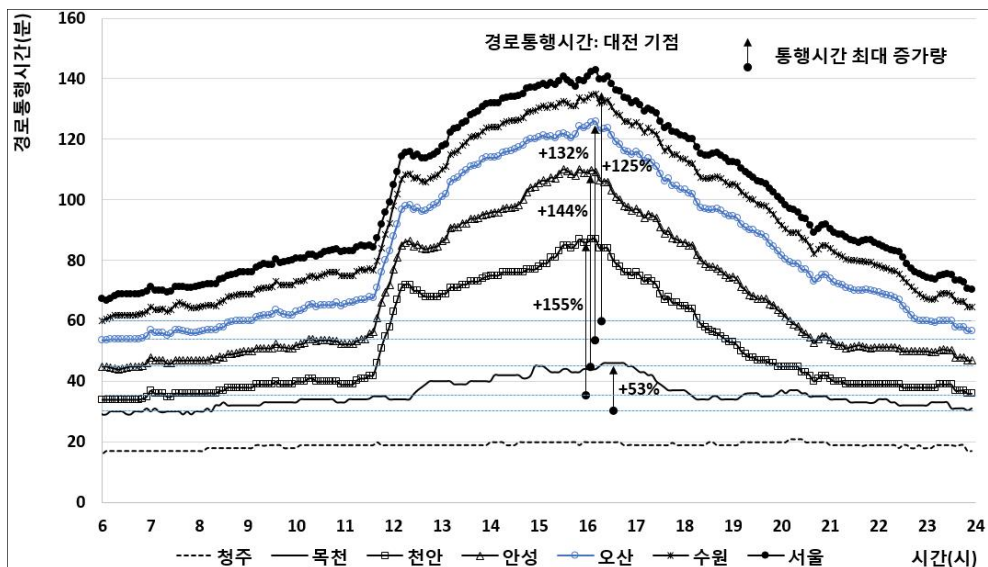
구간 프로브 통행량은 도로망 소통상태를 설명해주는 주요한 변수이며, <그림 4-42>는 프로브 통행량의 시계열 변화를 보여주고 있다. 구간 13의 경우 천안JC에서 경부와 천안-논산 고속도로의 통행량이 합류하면서 프로브 대수는 급격히 증가하고 있다. 84시퀀스 이전을 제외하면 대부분의 시간대에서 최소 프로브 대수(n_{min})를 만족하고 있다. 그리고 낮은 밀도상태에서 정상교통류는 Liu et al.(2008)이 지정한 바와 같이 시간대의 길이를 조정할 수 있으며, 본 연구의 경우 <그림 4-10>의 이전 시간대 방향 집계시간 확장을 이용하여 최소 표본수를 충족하는데 무리가 없을 것으로 판단된다. 84→168시퀀스까지 통행량이 증가함에 따라 구간통행속도(kph)는 <그림 4-43>에서 120→100까지 서서히 감소하고 있으며, 교통와해 이후 통행량은 최대 70%까지 감소하고 있다. 70%의 통행량 감소율은 Kerner(2002)가 관찰한 용량대비 포화교통량 비율인 66.7%($\approx 1/1.5$)와 유사한 결과이다. 이후 혼잡이 완화됨에 따라 프로브 통행량은 혼잡발생 이전의 상태로 회복되고 있다. 이상의 구간 프로브 통행량과 구간통행속도의 관계는 전형적인 교통량-속도의 관계를 보이고 있다[참고 <그림 3-9>].



<그림 4-42> 구간 프로브 통행량 구축자료의 특성

9) 경로통행시간 구축자료의 특성

경로통행시간은 구간통행시간과 달리 이전 구간통행시간의 증가분이 순차적으로 누적됨으로 경로상의 혼잡 발생시 경로통행시간은 급격히 증가하며, 혼잡의 완화시 경로통행시간은 완만하게 감소한다. 그리고 경로통행시간의 증감 기울기는 통행거리에 반비례하며 낮아진다. <그림 4-43>은 통행거리에 따른 경로통행시간의 변화량을 보여주고 있다. 혼잡 발생시 경로통행시간은 혼잡 발생시 통행거리가 증가함에 따라 53~155%까지 증가하고 있으며, 절대적인 경로통행시간의 양은 15~80분까지 증가하고 있다. 대전IC→서울TG간의 경로통행시간은 145분까지 증가하고 있으며, 이는 예측의 관점에서 볼 때 예측된 구간통행시간을 이용하여 경로통행시간을 추정할 경우 5분 시간대 길이를 기준으로 장래 29(=145/5)개의 시간대에 대한 구간통행시간 예측이 필요함을 의미한다. 추가로 시간대 $t+1$ 의 구간통행시간을 예측한 후 동시경로통행시간(IPTT)을 적용하여 경로통행시간을 산정할 경우, 경로통행시간의 증가에 비례하여 상태회복 시간이 증가함으로 예측오차는 증폭되며, 속도 Profile 등의 이력평균을 적용할 경우 통행시간 증폭을 설명할 수 없다.

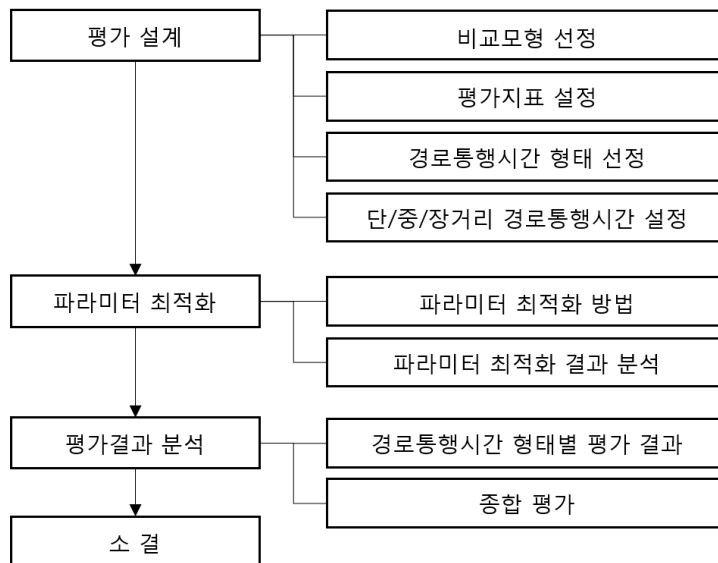


<그림 4-43> 경로통행시간 구축자료의 특성

V. 개발모형의 평가

1. 평가 개요

5장에서는 4장에서 구축된 자료를 이용하여 3장에서 개발된 KJC 예측기의 성능평가를 <그림 5-1>과 같이 수행하였다. 평가과정은 평가설계, 개발모형의 파라미터 최적화, 평가결과 분석 그리고 소결로 구성된다. 평가설계는 가장 널리 적용되는 비교모형의 설정, 비교모형과 개발모형의 성능평가를 위한 평가지표의 설정, 다양한 경로통행시간의 실질적인 평가를 위한 경로통행시간의 형태 선정과 단·중·장거리 통행시간의 설정으로 구성된다. 파라미터 최적화는 파라미터 최적화 방법을 설정한 후 파라미터 최적화와 결과 분석을 수행하였다. 그리고 파라미터 최적화 분석 결과를 기반으로 최적 입력변수를 선정하였다. 평가결과 분석은 설정된 평가지표와 선정된 경로통행시간 형태를 종합적으로 평가한 후, 개별 경로통행시간 형태에 대한 다각적인 미시적 평가를 수행하였다. 마지막으로 소결에서는 개발모형의 목표 달성 정도와 적용성에 대해 종합평가 결과에 근거하여 논하도록 한다.



<그림 5-1> 개발모형의 평가 과정

2. 평가 설계

1) 비교모형의 선정

개발모형은 도로망의 소통상태를 이용하여 단·중·장거리 통행시간을의 예측하게 된다. 따라서 기존의 경로통행시간 예측모형을 선정한 후, 비교모형의 예측성능을 기준점(Benchmark)으로 개발모형의 예측성능을 평가하였다. 비교모형은 가장 널리 적용되는 경로통행시간 산정모형으로 순간경로통행시간(Instantaneous path travel time, IPTT)과 경험기반 경로통행시간(Experience-based path travel time, EPTT)을 선정하였다. IPTT는 예측된 시간대 $t+1$ 의 상태가 장래에도 지속적으로 유지된다는 가정으로 경로통행시간을 산출하며, EPTT는 시간대 $t+1$ 이후의 장래 상태는 과거 통행시간 패턴을 따른다는 가정으로 경로통행시간을 산출한다. IPTT와 EPTT의 예측영역에 대한 가정과 경로통행시간 궤적은 <그림 3-27>과 같으며, 식(3-31)과 식(3-32)로 각각 산정된다.

본 연구에서 비교모형으로 선정된 IPTT와 EPTT의 시간대 $t+1$ 의 구간통행시간 예측값은 실제 관측값을 적용한다. 이는 IPTT와 EPTT의 적용에 있어 기존의 단기 구간통행시간 예측모형의 예측영역 한계라는 단점을 극복하고 100%의 예측 정확도를 전제로한 것이다. <그림 4-40>을 통해 살펴본 바와같이 구간검지체계에서 구간통행시간 자료의 수집에 필요한 시간처짐(τ)은 식(3-2)와 같이 구간통행시간의 증가에 비례한다. 예를 들어 5km 구간의 평균 통행속도가 10km/시일 경우 30분의 시간처짐이 발생하며, 시간대의 길이를 5분으로 적용하면 시간처짐의 개수는 $7(=30/5+1)$ 이다. 이는 단기 구간통행시간 예측모형의 예측영역이 장래 7개 시간대 즉, 장래 35분까지 확장되어야 시간대 $t+1$ 의 구간통행시간을 예측할 수 있음을 의미한다. 기존연구의 고찰 결과를 통해 살펴본 바와같이, 기존 구간통행시간 예측모형의 예측시간 영역은 이에 미치지 못하고 있다. 따라서 비교모형인 IPTT와 EPTT에 관측값을 예측된 구간통행시간 대신 적용하는 경쟁우위(Competitive advantage)을 부여하고, 본

연구의 경로통행시간 산출모형인 동적 구간기반 경로통행시간(Dynamic link-based path travel time, DLPTT)와 경로기반 경로통행시간(Path-based path travel time, PPTT)에 불리한 조건(Penalty)을 부여하는 것과 동일하다. 추가로 본 연구에서 적용되는 DLPTT와 PPTT의 시·공간 궤적과 예측영역은 <그림 3-28>과 같으며, 식(3-33)과 식(3-34)를 이용하여 각각 산정된다.

2) 평가지표의 설정

평가지표는 평균 절대 백분위 오차(Mean absolute percentage error, 이하 MAPE(%))와 적중률(%)을 이용하도록 한다. 그리고 미시적인 상태의 변동분석을 통해 예측모형의 완전 예측성(Perfect predictability)을 분석하도록 하며, 평가지표의 설명은 다음과 같다.

경로통행시간은 <그림 4-43>을 이용하여 설명한 바와 같이 통행거리와 혼잡의 강도에 따라 증가한다. 예를 들어 서울-대전간 경로통행시간은 비혼잡시 70분에서 혼잡시 240분 이상까지 증가한다. 이와 같이 상태의 변동폭이 큰 경우, MAPE(%)는 직관적이면서 공정한 평가의 기준을 제공해 주며(Smith et al., 2002), 식(5-1)과 같다. 여기서, x_i 는 관측값, \hat{x}_i 는 예측값, 그리고 n 은 전체 자료의 개수이다. 추가로 본 연구에서 개발된 예측모형의 파라미터 최적화 분석에 MAPE를 평가지표로 이용하도록 한다.

$$MAPE(\%) = \frac{\sum_{i=1}^n (|x_i - \hat{x}_i|/x_i) \times 100}{n} \quad \text{식(5-1)}$$

MAPE와 예측오차가 요구수준을 만족하면, 예측모형은 안정적인 예측력을 보여야 한다. 예측의 안정성은 적중률(%)을 이용하여 분석이 가능하다. 국내 네비게이션 경로정보제공서비스의 예측 정확도 목표는 연속류의 경우 평균 $\pm 10\%$, 단속류의 경우 평균 $\pm 15 \sim 20\%$ 정도로 설정하고

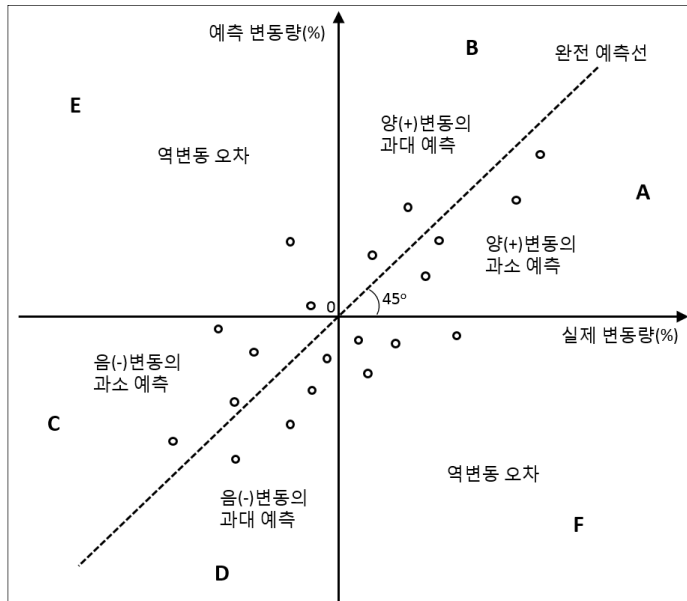
있다. 따라서 $\pm 10\%$ 이내의 적중률(%)은 예측모형의 안정성과 더불어 정확도 목표를 설명할 수 있는 효과적인 지표이다.

예측은 상태의 변동에 대한 장래의 상태를 설명하는 일련의 과정이라 할 수 있다. 즉, 상태의 방향성(Directionality)과 변화량(Variation)을 이용하여 상태의 진화를 설명하려는 일련의 노력이다. 따라서 관측치와 예측치의 변동분석을 통한 미시적 평가는 예측모형의 동적 특성 평가에 매우 중요하다. 변동분석은 관측값을 기준으로 실제 변동량(%)과 예측 변동량(%)의 관계를 분석하는 방법으로써 실제 변동량과 예측 변동량은 각각 식(5-2)와 식(5-3)과 같다. 여기서, $x(t)$ 는 시간대 t 의 관측값, $x(t+1)$ 은 시간대 $t+1$ 의 관측값, 그리고 $\hat{x}(t+1)$ 은 시간대 $t+1$ 의 예측값이다.

$$\text{실제 변동량(}\%) = 100 \times \frac{x(t+1) - x(t)}{x(t)} \quad \text{식(5-2)}$$

$$\text{예측 변동량(}\%) = 100 \times \frac{\hat{x}(t+1) - x(t)}{x(t)} \quad \text{식(5-3)}$$

식(5-2)와 식(5-3)의 변동량은 변동그래프를 이용하여 분석하게 되며, 변동그래프는 오차의 형태와 예측성을 직관적으로 보여준다. <그림 5-2>에서 $+45^\circ$ 기울기를 갖는 점선은 완전 예측선(Perfect forecast line)이며, 완전 예측선상의 점들은 관측값과 예측값이 완벽히 일치함을 의미한다. 변동분석 그래프에서 예측오차의 형태는 6개로 구분된다. A, B영역은 실제 변동량과 예측 변동량 모두 양(+)이지만, A영역은 과소 예측 B영역은 과대 예측을 의미한다. C, D영역은 실제 변동량과 예측 변동량이 모두 음(-)이지만, C영역은 과소 예측 D영역은 과대 예측의 상태이다. 따라서 A, B, C, D 영역의 경우 방향성은 예측되었으나 변화량의 예측에 오차가 발생한 경우이다. E와 F영역은 실제 방향과 예측 방향이 반대인 경우로서 방향성을 상실한 경우이다. 추가로 $\pm 5\%$ 정도 이내의 역변동은 상태의 자연적 변동으로 인한 오차임으로 역변동과는 구분된다.

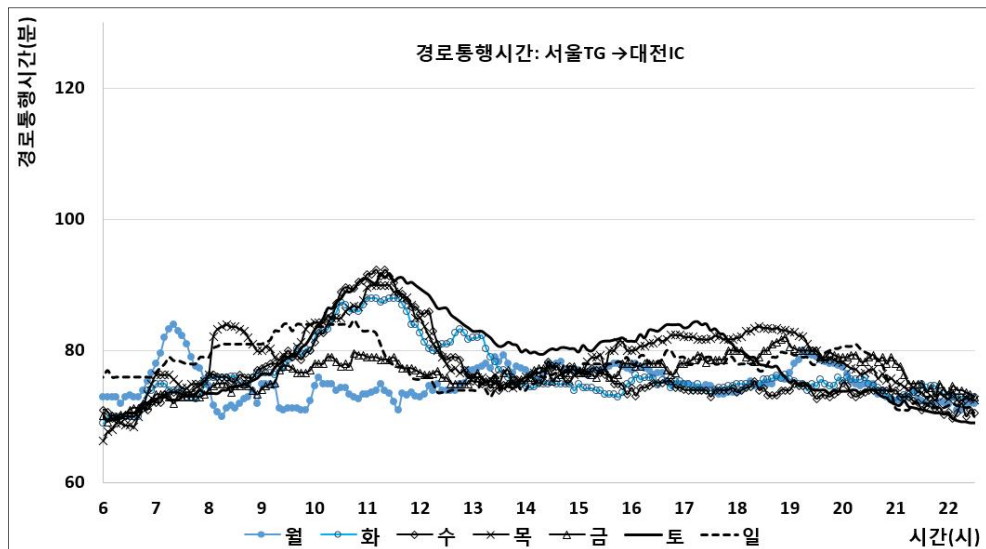


<그림 5-2> 변동분석 그래프

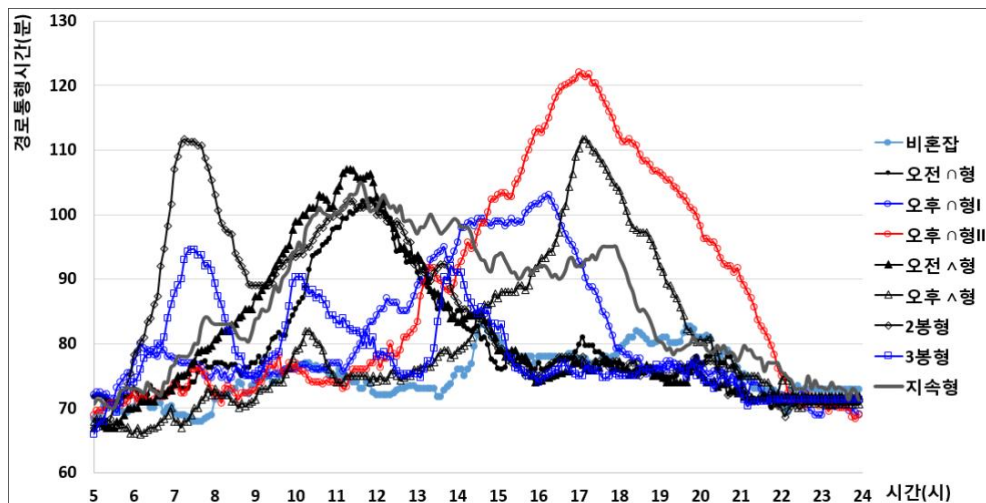
3) 경로통행시간 형태의 선정

고속도로 소통상태는 일주일간 혼잡이 발생하지 않은 경우와 일주일 동안 지속적인 주간 혼잡이 발생하는 경우 등으로 다양하게 변화하며, 이는 계절성, 날씨, 공사, 이벤트 등 다양한 원인 때문이다. <그림 5-3>은 2011년 03월 21일 ~ 27일까지 일주일 동안 서울TG→대전IC간 경로통행시간을 보여주고 있다. 토요일을 제외하면 나머지 요일의 경로통행시간은 평균 $\pm 10\%$ 이내에서 크게 변화하고 있지 않다. 다시말해 경로통행시간의 형태는 다양성이라는 측면에서 큰 차이를 보이지 않고 있다. 그러나 서울-대전간 통행시간은 실제로 70~300분까지 크게 변화한다. 따라서 특정 일주일을 선택하여 예측모형을 평가할 경우, 다양한 경로통행시간 행태에 대한 예측모형의 변별력을 평가할 수 없다. 또한 경로통행시간의 패턴이 단조로운 일주일을 선택하여 성능평가를 수행할 경우, 모형의 예측오차는 감소하겠지만, 실제 모형의 적용성은 현저히 감소한다. 특히, 장거리 통행시간의 경우 통행 중에 교통상황은 다양하게 변화한다. 따라서 경로통행시간은 다조로운 형태가 아닌 다양한 형태에서 평가되어야 한다.

따라서 본 연구에는 다양한 경로통행시간의 형태를 선정하여 예측모형의 수행성능을 비교·평가하였다. 선별된 경로통행시간의 형태는 서울-대전간을 기준으로 ①비혼잡, ②오전 ㄱ형, ③오후 ㄱ형I, ④오후 ㄱ형II, ⑤오전 ㄴ형, ⑥오후 ㄴ형, ⑦2 봉형, ⑧3 봉형, ⑨주간 지속형, ⑩극심한 정체로 총 10개가지이며, <그림 5-4>는 ①~⑨까지의 형태를 보여주고 있다.



<그림 5-3> 단조로운 경로통행시간의 형태



<그림 5-4> 선별된 경로통행시간의 형태

4) 평가를 위한 단·중·장거리 통행의 설정

본 연구에서 개발된 예측모형은 예측시점에서 장래 6시간까지 도로망 소통상태를 예측하고, 단·중·장거리의 경로통행시간을 모두 예측한다. 그러므로 통행거리를 단·중·장거리로 구분하여 경로통행시간을 평가하였다. 통행거리는 단거리 40km이하, 중거리 40~80km, 장거리 80km 이상으로 설정하였으며, <표 5-1>과 같다. 단거리는 서울TG를 기점으로 안성IC 이전이며 <그림 4-21>에서 경로통행시간 프로브 구성비가 20%까지 수집되는 거리이다. 중거리는 천안JC 이남의 목천IC까지 경로통행시간 프로브 구성비가 5%까지 수집되는 거리이다. 마지막으로 장거리는 목천IC 이남부터 대전IC까지이다.

<표 5-1> 단·중·장거리 통행 설정 결과

OD	대전	청주	목천	천안	안성	오산	수원	서울
대전	-	1	2	2	3	3	3	3
청주	1	-	1	1	2	2	3	3
목천	2	1	-	1	1	2	2	2
천안	2	1	1	-	1	1	2	2
안성	3	2	1	1	-	1	2	2
오산	3	2	2	1	1	-	1	1
수원	3	3	2	2	1	1	-	1
서울	3	3	2	2	2	1	1	-

주) 1=단거리, 2=중거리, 3=장거리 통행임.

3. 파라미터 최적화

1) 파라미터 최적화 방법

일반적으로 상태벡터를 이용하는 KNN기반의 예측모형은 상태를 재구성(Reconstruction)하기 위한 상태의 Embedding size 즉, 시계열 자료의 시간대 개수(l)와 k 개 이웃을 결정하는 k 값에 대한 최적값을 정산해야 한다. KNN-NPR 방법론에서 적정 파라미터를 최적화하기 위한 명확한 방

법론은 정립되지 않은 실정이다(Chang et al., 2012_b). 따라서 본 연구에서는 기존의 KNN-NPR 기반의 예측모형에서 통상 이용되는 데이터 시물레이션 기법을 이용하였다(Smith et al., 2002; Turochy, 2006; Chang et al, 2010; Chang et al., 2012_{a,b}; Yoon and Chang, 2014). 데이터 시물레이션 기법 최적화는 실제 관측값을 알고 있을 경우, 두 파라미터 값의 공간을 한정하고 두 파라미터 값의 모든 가능한 조합의 시나리에 대한 예측오차의 변화량 분석, 즉 민감도 분석을 수행함으로써 최적 오차공간에서 최적 파라미터 값을 결정하는 방법이다. 그리고 파라미터 자동 정산은 본 연구에서 개발된 KJC 예측기의 목표중 하나이다. 기존의 KNN-NPR 파라미터 최적화는 전역 오차 최소화를 목표로 최적 파라미터 값을 결정하였다. 따라서 예측의 정확도를 저하시킬 뿐만 아니라 주기적으로 파라미터 값을 갱신해야 하는 문제점이 있다. 본 연구에서는 적정 파라미터 값의 공간을 포함하도록 k 값을 크게 설정한 후 j -clustering 모형을 이용하여 국부 최적값을 결정함으로써 모형의 예측력을 최대한 향상시킴과 동시에 파라미터 최적화의 문제를 해결하고자 하였다. 추가적으로 적정 l 값은 한번 최적화되면 지속적으로 적용할 수 있으므로 모형의 예측력을 좌·우하는 파라미터는 k 값이 된다.

파라미터 최적화를 위한 입력변수의 시나리오는 S1=[구간통행속도, 구간 프로브 통행량], S2=[구간통행속도, 구간 프로브 통행량, 유·출입 통행량]이다. 최적화는 총 55,987,200 [=2개 입력변수 시나리오 × 10개 경로통행시간 형태 × 216(6~24시)개 시간대 × 36개 상태벡터의 시계열 개수(l) × 30개 k 값 × 12개 장거리 통행]건의 관측값을 각각 예측하고, MAPE(%)를 이용하여 이를 평가하였다. 전일의 자료를 이용하지 않고 6~24시간으로 한정 한 이유는 0~6시에는 일반적으로 혼잡이 발생하지 않기 때문에 오차공간의 오차를 줄이고 예측모형의 정확도를 향상시키게 되며, 이는 실제 통행시간이 증가하는 시간대의 증분된 예측오차를 희석시키게 된다. 다시말해, 실제 통행시간정보의 가치가 높은 시간대의 예측 정확도를 저해하는 원인이 되기 때문이다.

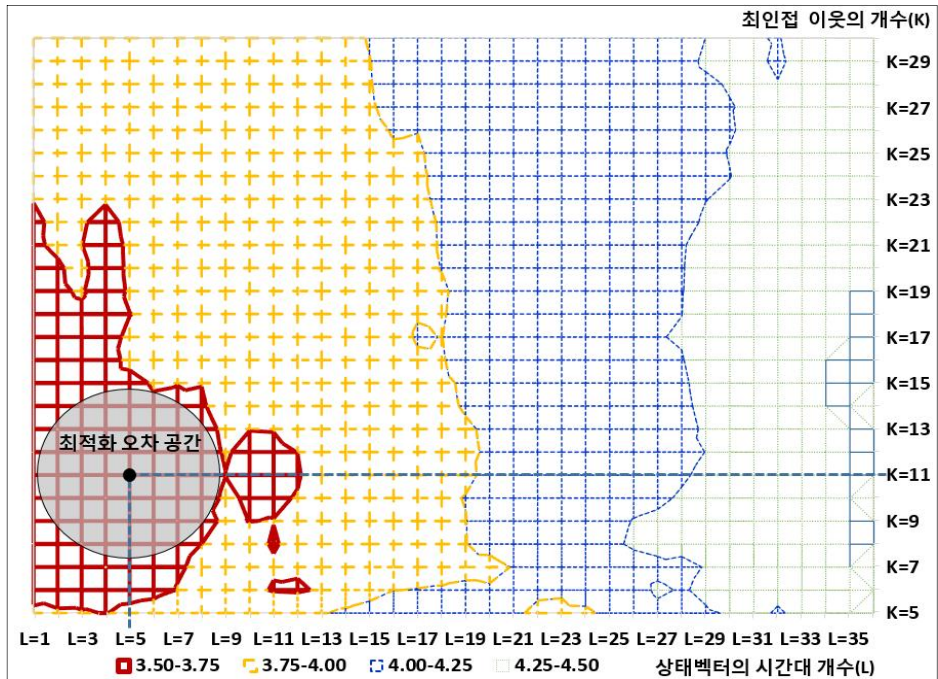
2) 파라미터 최적화 결과

<그림 5-5>와 <그림 5-6>은 2개 입력자료 시나리오에 대한 파라미터 최적화 결과를 보여주고 있다. S1[구간통행속도, 구간 프로브 통행량]의 오차공간은 S2[구간통행속도, 구간 프로브 통행량, 유·출입 통행량]의 오차공간보다 평균 0.25% 낮은 오차를 보이고 있다. 이는 구간통행속도와 구간 프로브 통행량의 장래 상태에 대한 설명력이 방향성을 고려할 수 없는 유·출입 통행량에 비해 높다는 직접적인 증거이다.

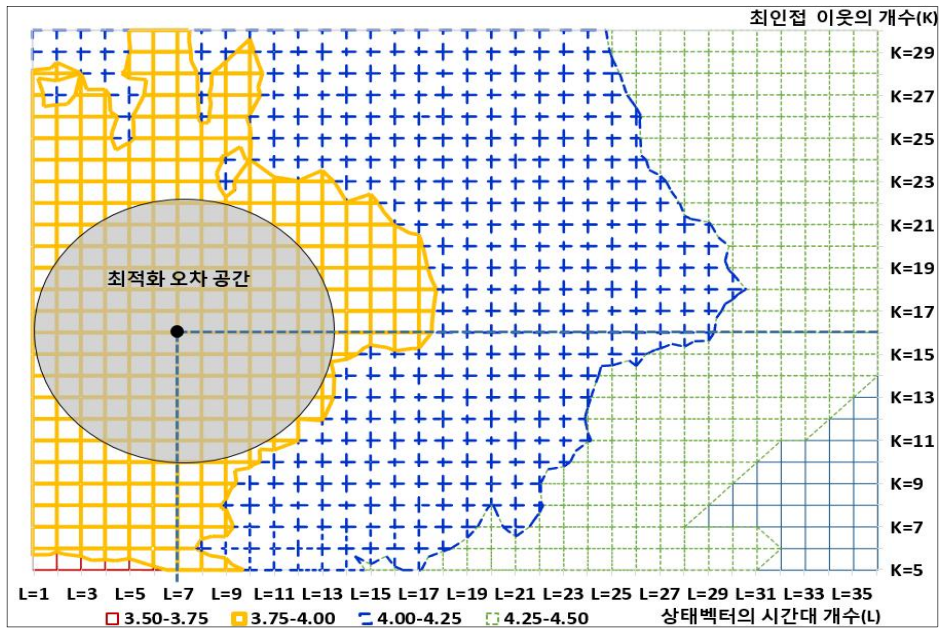
반면, 예측오차 3.5~4.0(%)의 수준의 경우, 두 입력변수 시나리오의 오차공간은 유사하게 나타나고 있다. 추가적으로 오차 0.25%를 감안할 경우, 두 입력변수 시나리오의 오차공간은 매우 안정적이다. 따라서 다양한 원인에 의한 불안정한 입력자료의 상태를 대비하여 두 입력변수 시나리오는 병렬적으로 이용할 수 있을 것으로 판단된다.

파라미터 최적값은 S1의 경우 $l=5$, $k=11$ 로, S2의 경우 $l=7$, $k=16$ 으로 분석되었다. S2의 파라미터 값이 상대적으로 큰 이유는 유·출입 통행량의 시계열 변동이 구간통행속도나 구간 프로브 통행량에 비하여 높게 나타나 상태구성을 위한 상태벡터의 Embedding size(l)를 최적화 과정에서 증가시켰기 때문이다. 다시말해, 유·출입 통행량의 경우 방향성을 알 수 없기 때문에 예측오차를 최소화시키는 과정에서 l 의 크기는 증가하였다. 추가적으로 S1의 경우 S2의 최적화 오차공간까지 확대할 경우 매우 안정적인 오차공간을 확보하고 있다. 이와 같이 본 연구에서 개발된 모형은 안정적인 오차공간을 확보하면서 오차를 감소시켰다.

이상의 파라미터 최적화 분석결과를 기반으로 개발모형의 평가에 적용되는 입력자료는 구간통행속도와 구간 프로브 통행량으로 선정하였다. 파라미터 값은 <그림 5-7>과 같이 결정하였으며, 예측오차를 최소화하는 파라미터 값의 조합은 $l=4$, $k=10$ 으로 나타나고 있다. 그러나 실제 개발모형의 적용을 고려하면 국부적인 오차 최소화로 판단하는 것이 바람직하지 않으며 적정 파라미터 값은 l 과 k 값의 특성분석을 통해 결정되어야 한다.

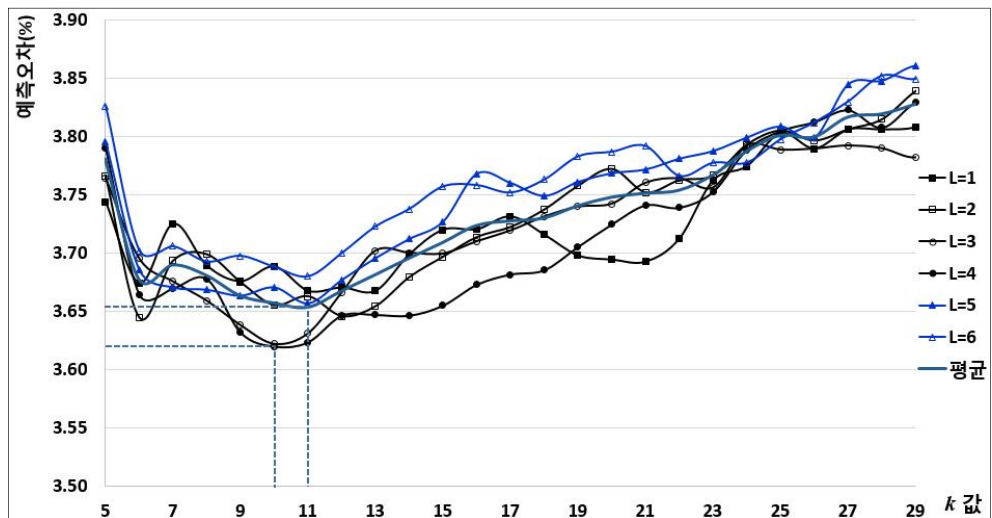


<그림 5-5> 파라미터 최적화 결과: 시나리오 1



<그림 5-6> 파라미터 최적화 결과: 시나리오 2

<그림 5-7>은 예측오차에 대한 l 과 k 값의 관계를 보여주고 있다. l 값이 증가함에 따라 예측오차는 감소하며, $l=4$ 에서 예측오차는 최소영역에 도달하고 있다. 이후 l 값이 증가하면서 예측오차는 다시 증가하고 있다. l 값에 의한 예측오차의 변화 영역은 0.1% 이내로 나타나고 있는 반면, k 값에 의한 예측오차의 변화 영역은 0.225% 이내로서 k 값의 파라미터 민감도는 l 값에 비하여 높게 나타나고 있다. 이는 네트워크 기반의 소통상태 예측에 있어 자료의 수집에 필연적으로 수반되는 시간처짐과 개별 도로구간의 교통상태 변화보다는 도로망의 전반적인 소통상태가 더 주요한 변수임을 간접적으로 말해준다. k 값이 증가함에 따라 오차는 급격히 최소 예측오차로 수렴해가며, 최소 오차에 도달한 이후 예측오차는 국부적인 임의 변동을 수반하면서 점진적으로 증가하고 있다. k 값을 기준으로 최적 오차의 공간은 예측오차 3.625~3.700의 범위내에서 $k=6\sim13$ 까지 나타나고 있다. 따라서 평균 예측오차를 기준으로 적정 파라미터의 값을 결정하면 $l_o=4$, $k_o=11$ 이다. 결정된 파라미터는 $l_o\pm2$, $k_o\pm5$ 의 범위내에서 예측오차 3.625~3.700의 확보가 가능하다.



<그림 5-7> 최적 파라미터 값의 선정

파라미터 최적화 결과, 입력자료는 구간통행속도와 구간 프로브 통행량의 조합이 유·출입 통행량을 추가한 조합보다 예측력이 우수한 것으

로 판단된다. 이는 개발모형의 적용이라는 관점에서 매우 주요한 장점이
다. 일반 도로망은 유·출입 통행량의 실시간 수집이 현장 여건상 용이
하지 않은 반면, GPS, 스마트 폰 등과 같은 스마트 개인단말을 통해 수
집되는 구간 통행속도와 구간 프로프 통행량은 도로망의 위계에 관계없
이 지속적으로 수집되기 때문이다.

장거리 경로통행시간 예측의 관점에서 l 과 k 의 모든 범위에서 예측오
차는 4% 이하로서 매우 낮으며, 예측오차의 요구수준을 충족하고 있다
고 판단된다. 이는 KJC 예측모형의 KN 단계에서 도로망의 소통상태를
이용하여 장래 상태를 설명하는 1차 군집을 구축한 후, j -군집화 과정을
통해 동적으로 최적의 의사결정 군집을 세분화하기 때문이다. 따라서
KJC 예측기의 개발시 주요 고려사항인 파라미터 최적화의 자동화 기능
과 더불어 수용 가능한 예측오차 공간의 확대라는 부분적인 목표는 달성
한 것으로 판단된다.

4. 평가 결과

1) 종합평가 결과

장·중·단거리 통행시간에 대한 종합평가 결과는 <표 5-2~4>와 같
다. 본 연구에서 개발된 KJC 예측기로 추정된 장래 소통상태를 이용하
여 생성된 경로통행시간인 DLPTT와 PPTT는 예측오차 $\pm 0.6\%$ 이내
에서 동일하다고 판단된다. 반면 과거의 평균 구간통행시간을 이용하는
EPTT는 예측오차 기준으로 가장 낮은 예측력을 보였다.

장거리 경로통행시간 평가결과, EPTT의 평균 예측오차는 12% 이상을
보이고 있다. 이는 지역간 고속도로의 경우 통행의 반복적 행태가 약하
기 때문으로 판단된다. 오전 시간대의 예측결과는 모든 모형에서 오후
시간대보다 우수하게 나타났다. 이는 오전 시간대 통행의 반복적 특성이
오후 시간대에 비하여 강하기 때문으로 판단된다. 따라서 과거의 평균
구간통행시간을 이용하는 EPTT의 오전 시간대 예측결과는 IPTT에 비
하여 우수하게 나타났으며, 이는 오전 시간대 통행의 반복적 특성이 강

하다는 직접적인 증거라고 할 수 있다. 반면 반복적 통행특성이 낮은 오후 8형의 경우, EPTT의 예측력은 가장 낮게 나타났다. 비혼잡, 11형, 8형에서 DLPTT와 PPTT의 예측오차는 4% 미만 적중률은 평균 95% 이상으로 IPTT와 EPTT에 비하여 매우 우수하게 나타났다. 이와같이 반복적/비반복적 통행의 모든 부분에서 KJC 예측기가 우수한 예측성능을 보인 이유는 강력한 패턴인식을 기반으로 개발되었기 때문이다. 2봉, 3봉, 주간 지속형의 경우 IPTT와 EPTT의 예측성능은 낮아진 반면, 개발모형의 성능은 일봉형과 큰 차이를 보이고 있지 않다. 추가로 주간 지속형의 경우 개발모형의 적중률은 98%에 가깝게 분석되었다. 극심한 정체성의 경우 IPTT와 EPTT는 10% 이상의 예측오차를 보인 반면 개발모형의 경우 8%이내를 유지하고 있다. 적중률은 70% 초반대로 매우 저하되고 있으나, IPTT와 EPTT의 적중률은 24.60~48.27%을 보이면서 완벽한 예측실패를 보였다. 극심한 정체상태에서 개발모형의 성능이 저하된 이유는 대전-서울간 경로통행시간이 280분에 달하는 이력자료의 다양성이 부족하기 때문이며, 이는 이력자료의 양을 확대함으로써 해결이 가능하다고 판단된다.

<표 5-2> 장거리 경로통행시간의 종합평가 결과

통행시간 형태	요일	예측오차(%)				±10이내 적중률(%)			
		IPTT	EPTT	DLPTT	PPTT	IPTT	EPTT	DLPTT	PPTT
비혼잡	일	2.69	47.81	2.66	2.49	98.25	47.81	99.85	99.92
오전 11형	화	3.52	2.65	2.23	2.32	96.57	98.69	99.85	99.56
오후 11형I	월	6.13	4.64	3.28	3.37	82.04	89.59	95.25	95.15
오후 11형II	금	7.80	5.56	3.96	3.84	63.32	81.15	93.15	92.35
오전 8형	수	3.57	3.87	2.98	2.23	98.32	92.07	99.20	99.41
오후 8형	일	7.37	15.77	3.72	3.57	76.49	43.01	95.92	95.78
2봉형	월	3.86	3.05	3.51	2.85	95.05	96.87	93.74	96.36
3봉형	월	4.64	6.81	3.51	3.04	87.55	79.99	94.15	93.86
주간 지속형	금	4.05	4.34	2.43	2.51	95.74	83.41	97.91	97.91
극심한 정체	일	13.01	30.43	7.31	7.46	48.47	24.60	72.76	72.02
평 균		5.66	12.49	3.56	3.37	84.18	73.72	94.18	94.23

중거리 통행시간 평가결과, 모든 모형의 예측오차는 장거리 통행시간에 비하여 증가하였으며, 적중률은 낮아졌다. 이는 절대적인 통행시간이 짧아지면서 상대적인 지표인 MAPE가 증가했기 때문으로 판단된다. 개발모형의 예측오차는 평균 5%이하, 적중률은 평균 90%이상을 유지하고 있으며, 이는 수용가능한 예측오차 범위인 10%내에서 개발모형의 예측안정성이 유지되고 있음을 말해준다. EPTT의 경우 비혼잡시와 오후 ⅸ형 예측오차는 10% 이상을 보이고 있지만 장거리 통행시간의 예측오차에 비하여 대폭 감소하였다. 이는 장거리 통행에 비하여 중거리 통행의 반복적 통행행태가 높다는 간접적인 의미이다. 극심한 정체와 경우 IPTT와 개발모형의 적중률은 $\pm 1\%$ 내에서 유사하게 나타나고 있다. 이는 심한 정체상태에서 교통류는 동질적 특성을 보이기 때문에 현재 상태가 장래에도 지속된다는 IPTT의 장래상태에 대한 가정을 간접적으로 충족시켰기 때문이다. 동일한 맥락에서 IPTT의 경우 오후 Ⅱ형의 예측오차는 장거리 통행시간에 비하여 감소하고 있다. 그러나 예측 정확도는 개발모형에 못 미치는 것으로 나타나고 있다.

<표 5-3> 중거리 경로통행시간의 종합평가 결과

통행시간 형태	요일	예측오차(%)				± 10 이내 적중률(%)			
		IPTT	EPTT	DLPTT	PPTT	IPTT	EPTT	DLPTT	PPTT
비혼잡	일	2.97	14.36	3.02	2.86	96.76	50.46	98.95	98.68
오전 Ⅱ형	화	4.16	3.32	2.65	2.58	92.44	95.55	99.39	99.34
오후 Ⅱ형Ⅰ	월	6.45	7.83	4.42	4.37	81.93	74.43	87.71	88.75
오후 Ⅱ형Ⅱ	금	7.76	7.57	4.64	4.52	67.13	70.62	86.54	86.42
오전 ⅸ형	수	4.28	5.13	3.64	2.93	90.87	86.86	96.94	97.68
오후 ⅸ형	일	6.99	12.56	3.77	3.74	75.90	56.70	93.61	93.18
2봉형	월	4.86	3.65	3.82	3.35	88.60	92.97	93.75	94.93
3봉형	월	5.89	7.57	4.18	3.99	81.70	76.42	89.74	89.21
주간 지속형	금	5.37	8.00	3.88	3.85	84.33	69.38	92.73	92.18
극심한 정체	일	9.55	24.24	8.75	8.56	66.61	32.95	65.80	66.80
평 균		5.83	9.42	4.28	4.08	82.63	70.63	90.52	90.72

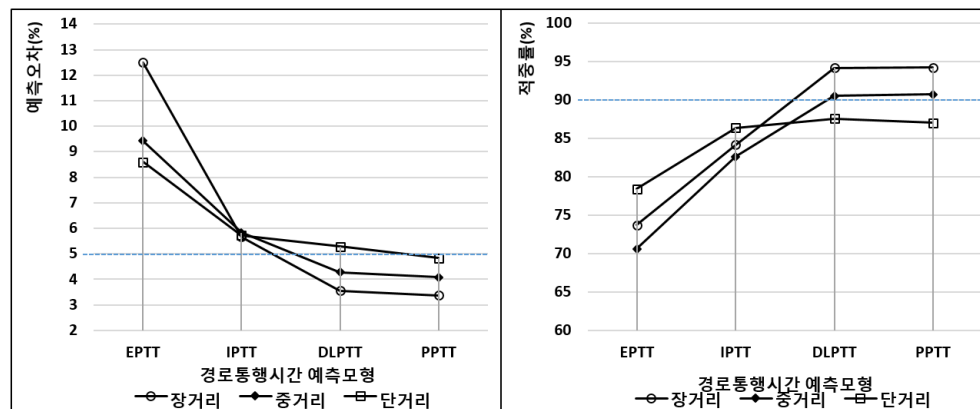
단거리 통행시간 평가결과, IPTT와 EPTT의 예측오차는 감소하고 적중률은 증가한 반면, 개발모형의 예측오차는 증가하고 적중률은 90%이하로 감소하였다. 그러나 개발모형의 예측오차는 5% 대를 유지하고 있으며, 적중률은 87% 이상으로 IPTT와 EPTT보다 높게 나타났다. EPTT의 경우 비혼잡시와 오후 ㄱ형의 예측오차는 10% 이내로 감소하였으며, 통행거리가 감소함에 따라 예측오차는 감소하고 있다. 이는 단거리 통행의 경우 반복적 통행시간 특성이 강한 도시부의 존간 경로통행시간이 반영되었기 때문이라고 판단된다. 극심한 정체를 제외하면 개발모형의 예측 성능은 IPTT와 EPTT에 비하여 높게 나타나고 있다. 추가로 극심한 정체의 경우 IPTT가 개발모형보다 높게 나타난 이유는 시간대 $t+1$ 의 실제 구간통행시간을 적용한 경쟁우위(Competitive advantage) 조건을 부여하였기 때문이다. 실제로 극심한 정체상황에서 구간검지체계에서 수집되는 구간통행시간 자료를 이용하기 위해서는 일반 혼잡보다 긴 시간처짐이 발생하기 때문에, 단기예측기법을 이용하여 예측된 구간통행시간을 이용하는 IPTT의 경우 예측치가 과거값이 되는 예측 실패(Forecasting failure)를 피할 수는 없다.

<표 5-4> 단거리 경로통행시간의 종합평가 결과

통행시간 형태	요일	예측오차(%)				±10이내 적중률(%)			
		IPTT	EPTT	DLPTT	PPTT	IPTT	EPTT	DLPTT	PPTT
비혼잡	일	3.75	9.08	3.58	3.27	95.81	69.97	96.40	94.69
오전 ㄱ형	화	4.78	4.50	3.78	3.15	90.75	91.01	96.46	95.39
오후 ㄱ형I	월	5.50	9.50	5.41	4.97	87.19	81.69	86.43	85.66
오후 ㄱ형II	금	5.82	7.00	4.41	4.93	84.48	78.92	86.22	86.58
오전 ㄴ형	수	4.66	5.16	4.21	3.60	88.25	87.85	92.00	92.32
오후 ㄴ형	일	5.70	11.13	4.70	4.35	87.34	72.37	90.37	88.84
2봉형	월	5.59	4.90	4.39	3.72	87.95	89.81	91.59	91.96
3봉형	월	6.59	7.37	5.02	4.52	85.33	83.52	88.89	88.01
주간 지속형	금	6.19	8.07	5.90	5.51	82.14	77.34	84.68	83.46
극심한 정체	일	8.76	19.16	11.41	10.40	74.83	52.04	63.03	63.59
평 균		5.73	8.59	5.28	4.84	86.41	78.45	87.61	87.05

<그림 5-8>은 통행거리에 따른 비교모형과 개발모형의 경로통행시간 예측성능을 종합하여 보여주고 있다. 장거리 통행시간의 경우 개발모형의 전체 평균 예측오차는 3.46%로서 IPTT대비 38.85%의 오차 감소를 보였으며, 적중률은 94.21%로서 IPTT대비 11.91% 높게 나타났다. 중거리 통행시간의 경우 개발모형의 전체 평균 예측오차는 4.18%로 나타나 IPTT대비 28.35%의 오차 감소를 보였으며, 적중률은 90.62%로서 IPTT 대비 9.67%의 향상을 보였다. 단거리 통행시간의 경우 개발모형의 전체 평균 예측오차는 5.06%로서 IPTT대비 11.73%의 오차 감소를 보였으며, 적중률은 87.33%로서 IPTT와 유사한 결과를 보였다.

이상의 결과를 종합하면 EPTT는 단·중·장거리 통행시간 예측에 있어 예측오차와 적중률을 고려할 때 요구수준을 충족하지 못 하였다. IPTT는 평균 6%로 안정적인 예측오차를 보이지만 적중률은 90% 이하로 나타났으며, 단거리 통행시간에 적합한 것으로 판단된다. DLPTT와 PPTT의 중·장거리 통행시간 예측오차는 5% 이하, 적중률은 90% 이상을 안정적으로 유지하였으며, 단거리 통행시간 예측오차는 5% 내외를 유지하면서 적중률은 87% 이상을 유지하였다. 따라서 본 연구에서 개발된 DLPTT와 PPTT는 장거리 뿐만 아니라 단·중거리 통행시간의 예측에도 기존 경로통행시간 예측모형보다 우수한 것으로 판단된다.



<그림 5-8> 통행거리에 따른 예측모형별 성능 비교 결과

DLPTT와 PPTT의 경우, 예측오차와 적중률의 관점에서 중·장거리 통행시간의 예측성능은 PPTT가 우수하게 나타났으며, 단거리 통행시간의 예측성능 중 예측오차는 PPTT가 적중률은 DLPTT가 우수하게 나타났다. 그러나 두 모형간의 차이는 크지 않은 것으로 판단된다. 따라서 모형의 적용성을 고려한다면 경로가 고정되지 않아 최적경로 탐색이 용이한 DLPTT가 PPTT보다 일반도로에 적합하다고 판단된다.

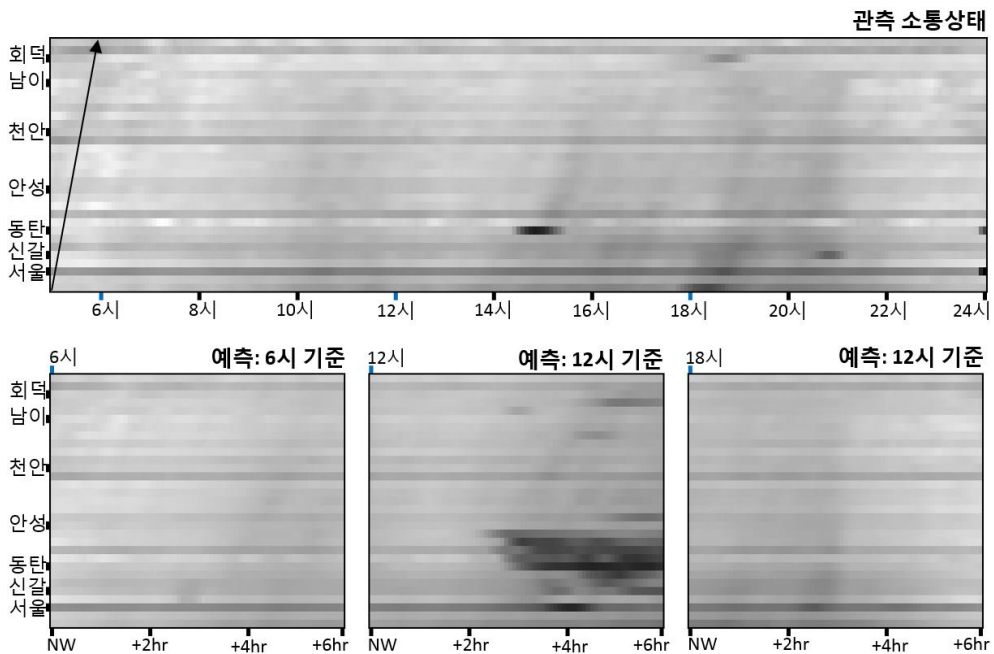
종합평가 결과, 통행거리가 증가함에 따라 비교모형과 개발모형의 예측력은 명확한 차이를 보였다. 그 원인으로는 개발모형의 경우 개별 도로구간이 아닌 도로망의 차원에서 소통상태를 예측함으로써 장래상태의 불확실성을 감소시켰기 때문이다. 비교모형의 경우 현재의 소통상태가 장래에도 지속된다는 가정이나 과거의 상태가 반복된다는 가정은 지역간 도로와 같이 통행행태가 다양한 도로망에서는 적합하지 않은 것으로 밝혀졌다. 본 연구에서 개발된 KJC 예측기는 이력자료에 내재된 다양한 행태에 대한 강력한 패턴인식을 기반으로 네트워크 기반의 소통상태를 예측한다. 따라서 소통상태의 예측에 있어 도로망에 내재된 복잡한 동적 행태를 이용하는 접근법은 장래 불확실성을 감소시킬 수 있는 방안 중 하나라는 직접적 증거이다.

2) 적용 결과: 비혼잡

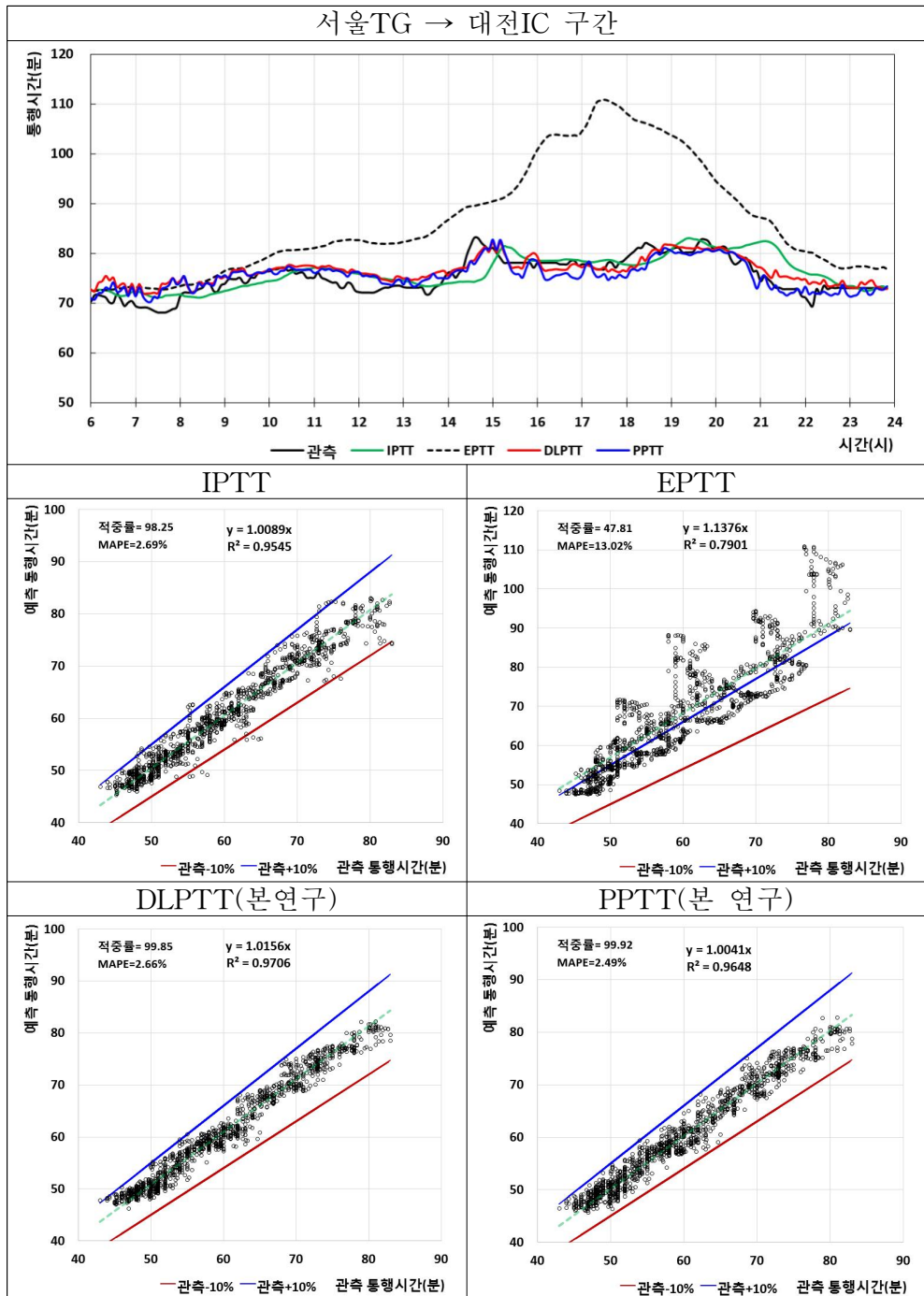
<그림 5-9>는 서울→대전방향의 31개 구간에 대한 6~24시까지 소통상태와 6, 12, 18시를 기준으로 장래 6시간까지 예측한 결과를 보여주고 있다(참고: 모든 시간대와 모든 구간의 속도 중 가장 낮은 관측속도에 가까울수록 검은색이며, 가장 높은 관측속도에 가까울수록 흰색이다). 6~24시까지 5분 간격으로 장래 6시간 동안 소통원할 상태를 각각 예측한 결과는 다음과 같다. 12시 기준으로 15시까지 비 혼잡상태를 예측하였으며, 15시의 동탄JC 부근에서 속도가 감소한 정상상태를 예측하였으나 공간과 시간을 다소 넓게 예측하였다(참고: 소통상태는 관측 구간속도에 대한 상대적인 값이므로 12시 기준으로 3시간 이후를 예측한 경우

는 혼잡상태가 아님). 전반적으로 1일 동안의 비혼잡상태를 수용할 만한 수준에서 예측했다고 판단된다.

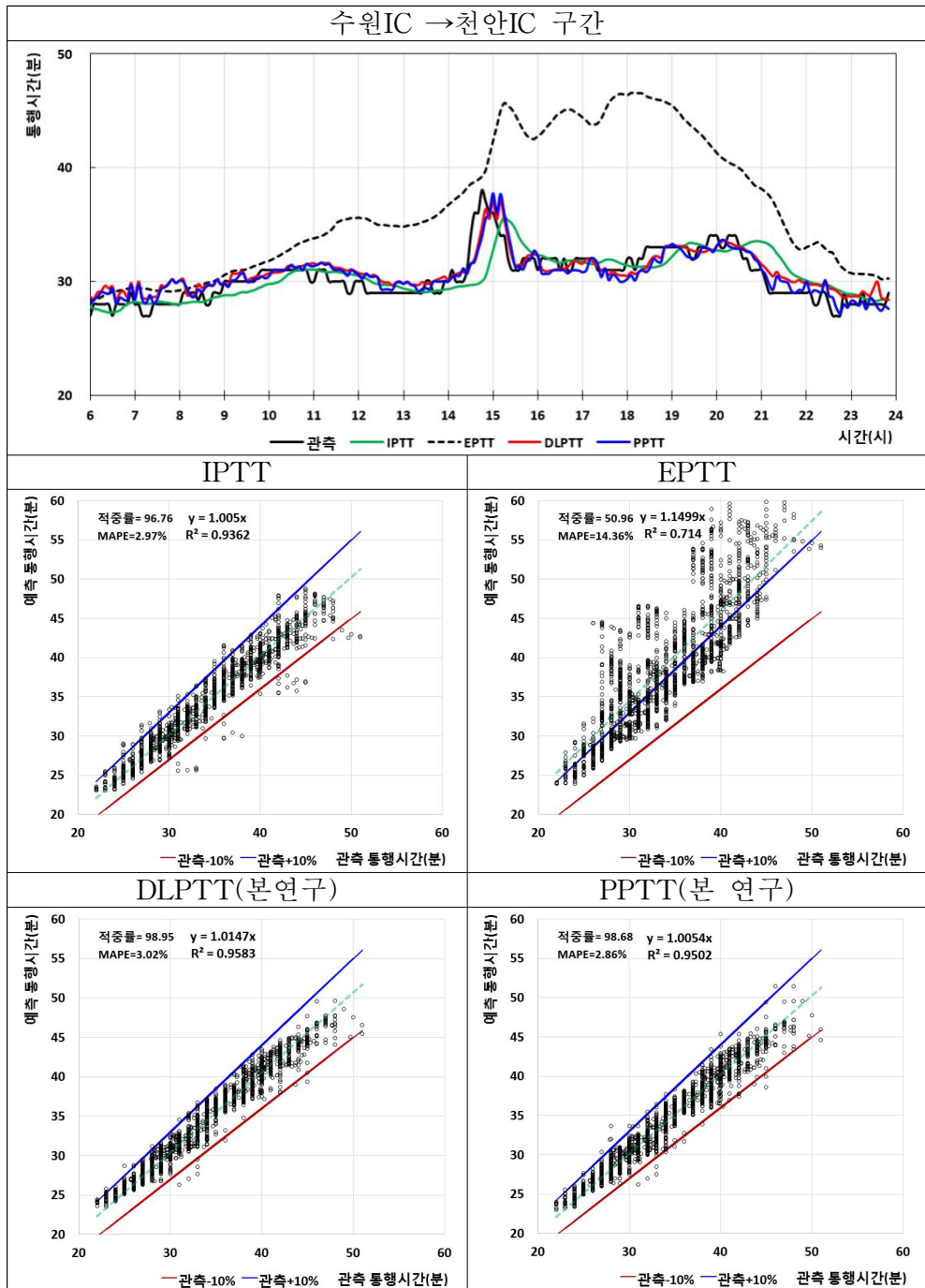
비 혼잡상태의 경로통행시간 행태는 Rolling 현상이 수반된다. 혼잡은 발생하지 않았지만 교통수요의 증가로 인한 속도감소 때문이며, <그림 5-10>에서 관측 통행시간은 전형적인 Rolling 현상을 보여주고 있다. IPTT의 경우 Rolling 현상으로 인해 발생한 낮은 침투에서도 상태회복 시간은 길게 나타나고 있다. EPTT의 경우 완전 예측실패를 보이고 있으며, 이는 대상일의 경로통행시간 행태가 비 반복적 특성이 강함을 의미한다. 반면, 개발모형은 Rolling 현상을 전반적으로 설명하고 있으며, 이는 구간 통행수요의 변동을 설명하고 있음을 의미한다. 따라서 장거리 통행시간의 경우 예측오차 2.49%이하에서 적중률은 99.85% 이상을 보이고 있다. 중거리 통행시간은 예측오차 3.02%이하에서 적중률 98.68% 이상, 단거리 통행시간은 예측오차 3.58%이하에서 적중률 94.69% 이상으로 나타났다.



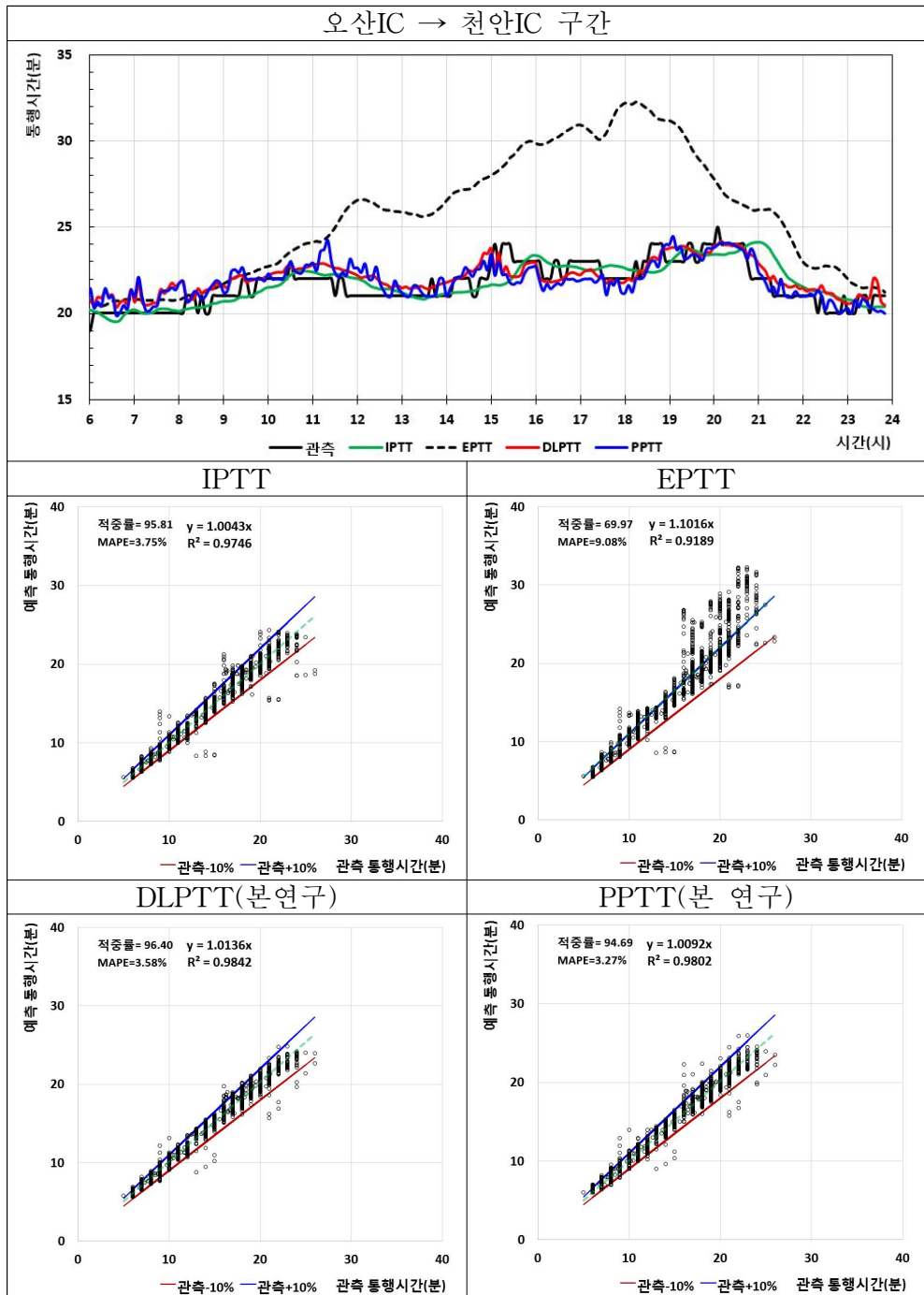
<그림 5-9> 소통상태 예측결과: 비혼잡



<그림 5-10> 장거리 경로통행시간 예측결과: 비혼잡

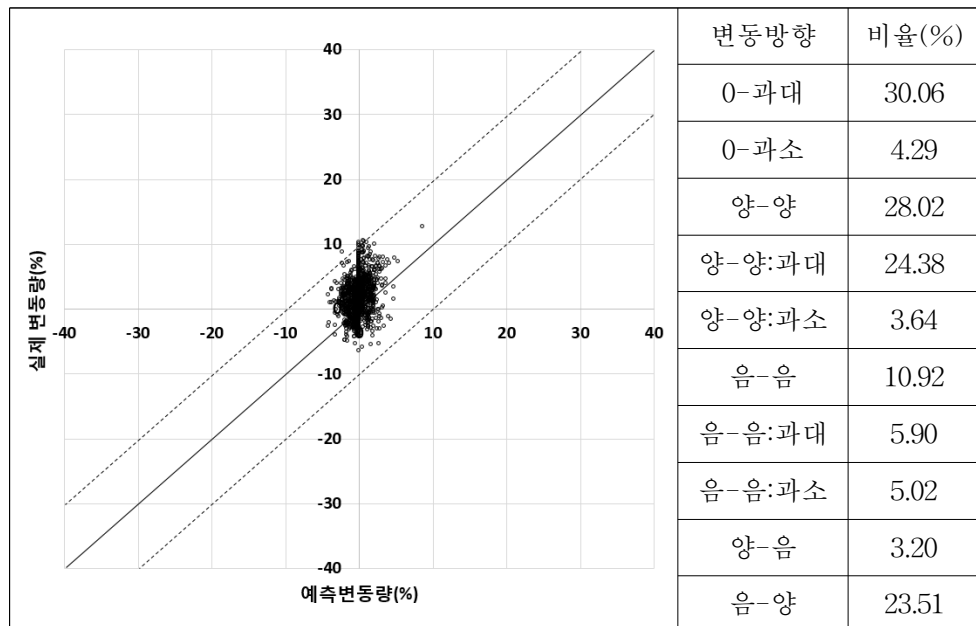


<그림 5-11> 중거리 경로통행시간 예측결과: 비혼잡

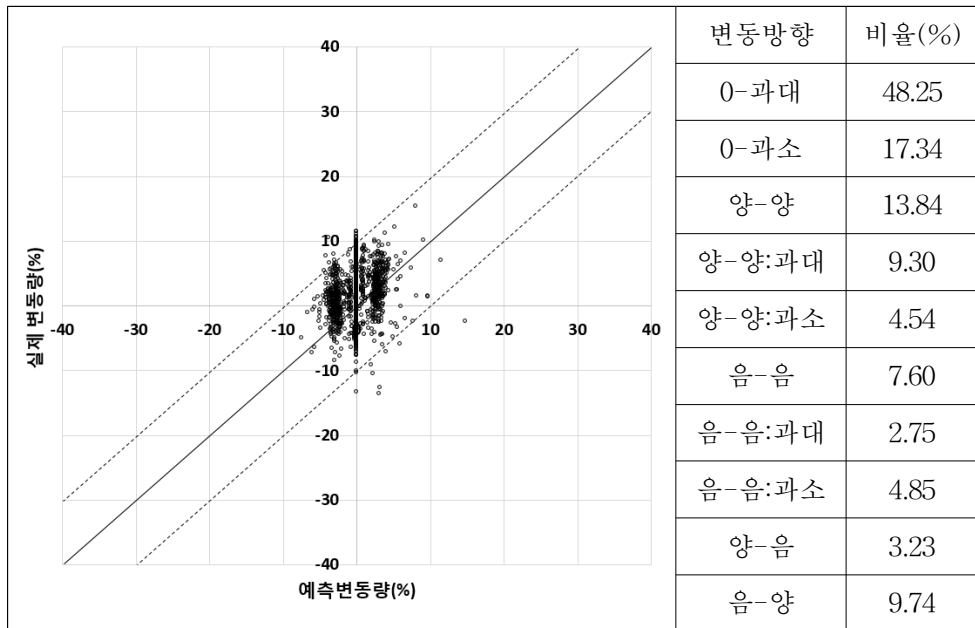


<그림 5-12> 단거리 경로통행시간 예측결과: 비혼잡

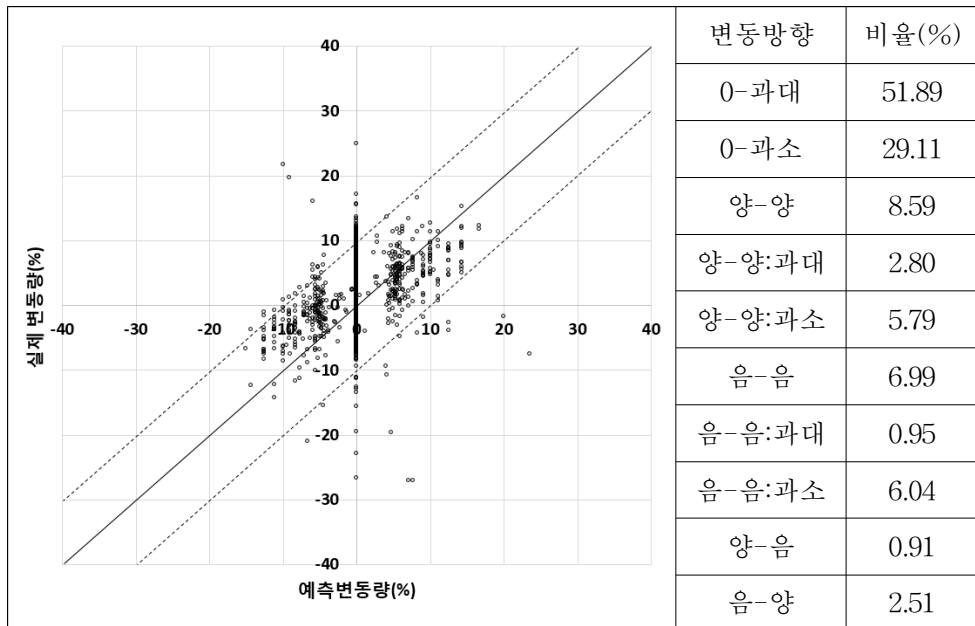
<그림 5-13~15>는 실제변동량(%)과 예측변동량(%)을 변동분석 그래프로 보여주고 있다(참고: 변동방향 X-Y는 실제변동량-예측변동량을 의미함). 참고로 통행시간의 경우 자체 변동을 수반하고 있으며, 자료의 수집·집계과정에서 발생하는 인위적인 변동이 포함되어 있다. 따라서 실제 통행시간의 예측에 있어 실제 차이가 크지 않다면 변동방향의 일치여부는 그렇게 주요하지 않다. 장거리 경로통행시간의 변동점은 중앙에 집중해있으며, 0-과대, 양-양, 음-양이 높게 나타나고 있어 예측치의 변동방향이 관측치의 변동방향을 설명하지 못 하고 있다. 그러나 변동량은 대부분 $\pm 5\%$ 이내에 집중되어 있기 때문에 예측결과는 우수하다고 판단되며, 예측치의 변동량 오차 $\pm 10\%$ 이내는 99.85%로서 상태의 방향을 크게 벗어나지 않고 있다. 단·중거리 통행시간의 경우, 0-과대, 0-과소가 높게 나타나고 있으며, 이는 비혼잡 상태에서 실제 통행시간 변동량이 크지 않기 때문이다. 반면 양-음, 음-양의 역변동은 낮게 나타나고 있으며 대부분 $\pm 10\%$ 이내에서 발생하고 있다.



<그림 5-13> 장거리 경로통행시간 예측 변동분석: 비혼잡



<그림 5-14> 중거리 경로통행시간 예측 변동분석: 비혼잡

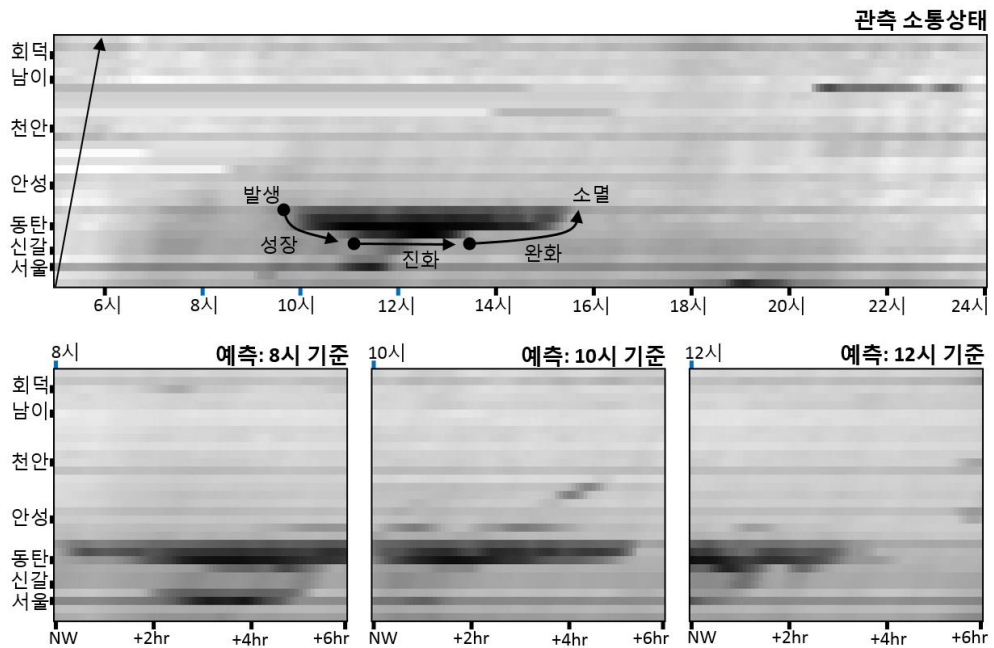


<그림 5-15> 단거리 경로통행시간 예측 변동분석: 비혼잡

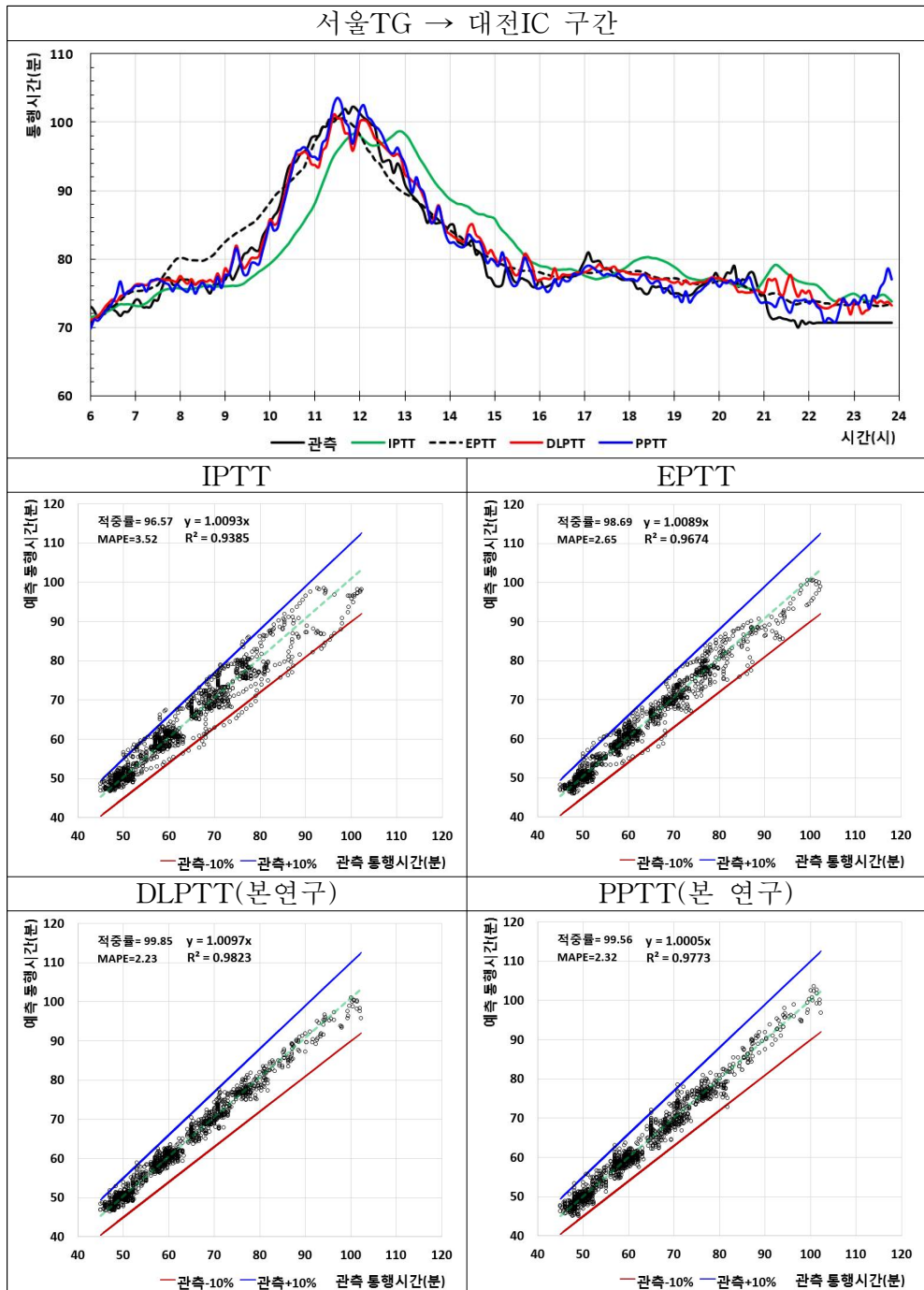
3) 적용 결과: 오전 \cap 형

\cap 형의 통행시간은 혼잡발생 이후 비교적 짧은 공간적 성장 후 일정 시간동안 시간적 진화과정은 유지되지만 공간적으로 진화하지 않는 전형적인 Stationary jam(SJ)에서 주로 나타난다. 예측모형의 경우, 예측시점 8시를 기준으로 2시간후 본격적인 혼잡을 예측함과 동시에 10~12시의 서울TG 구간의 혼잡을 예측하고 있다. 그리고 10시와 12시 예측시점에서 혼잡의 지속과 소멸을 예측하고 있다.

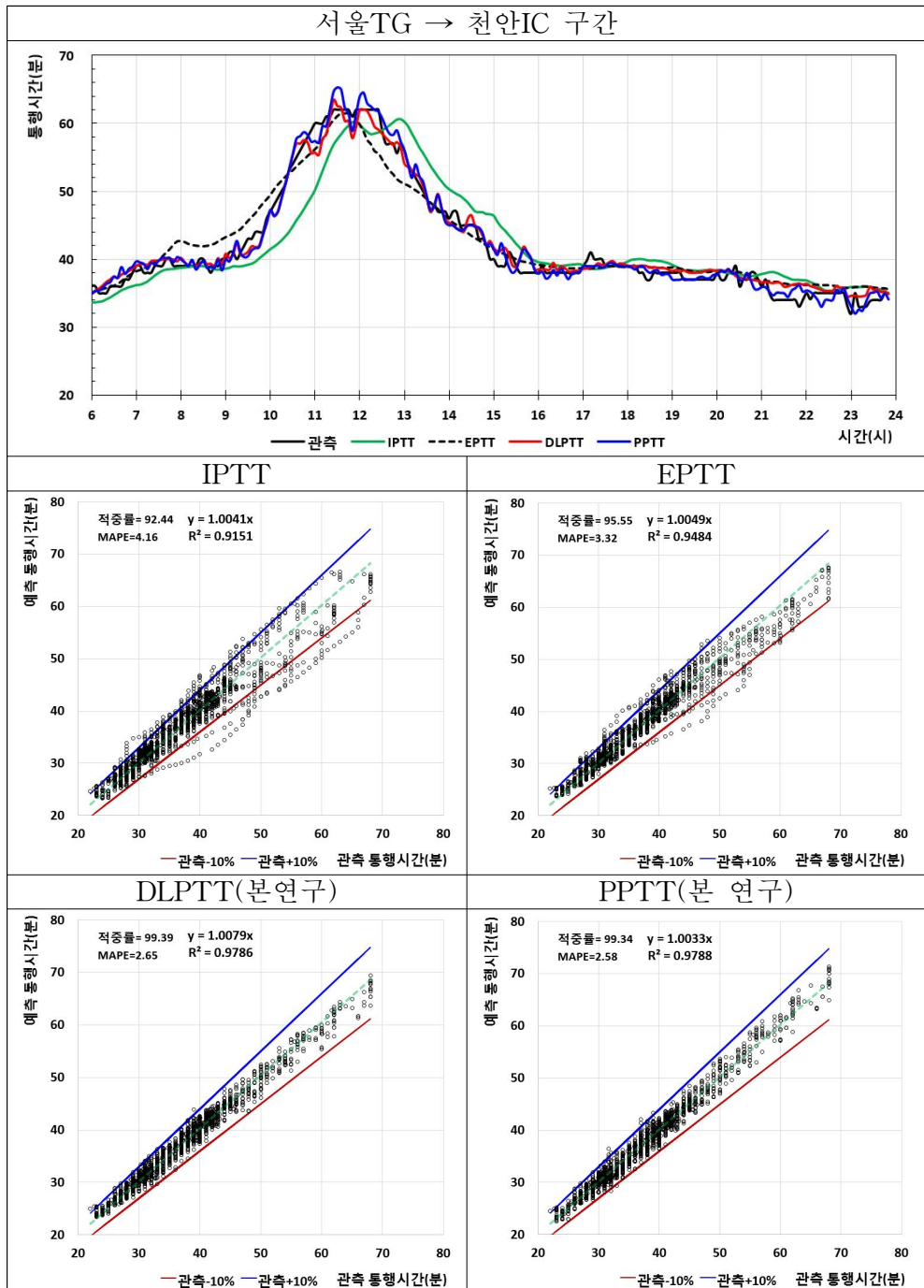
이상의 매우 정확한 6시간까지의 소통상태 예측으로 인하여 예측오차는 2%대로 진입하고 있으며, 적중률은 99.5%보다 높게 나타나고 있다. EPTT의 예측력도 적중률 98.69%로 나타나 큰 차이가 없는 것으로 나타나지만, 통행시간 80분 이상에서 개발모형의 예측력이 보다 우수한 것으로 나타나고 있다. 이는 개발모형이 반복적 행태에서도 국부적으로 내재된 패턴을 인식하기 때문이다.



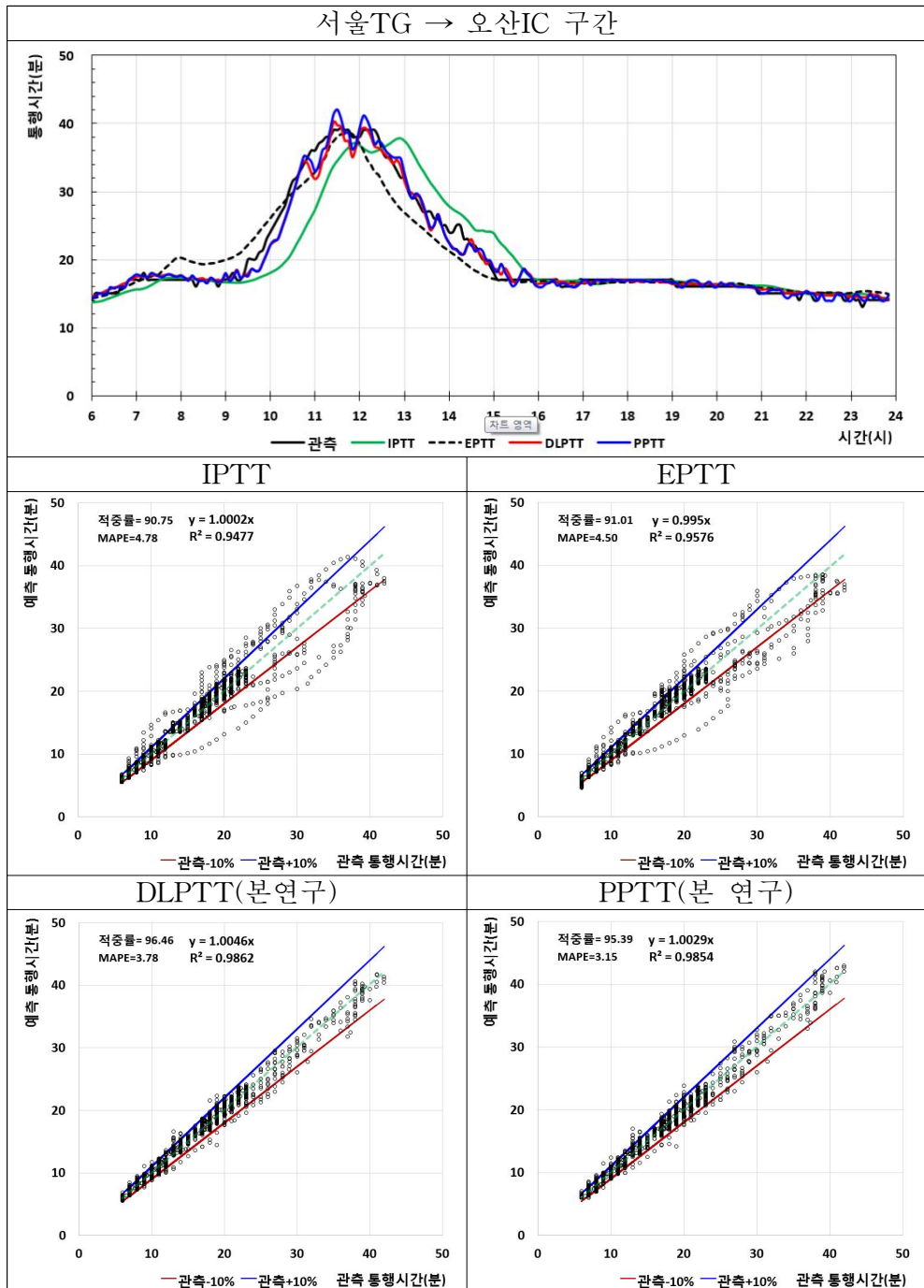
<그림 5-16> 소통상태 예측결과: 오전 \cap 형



<그림 5-17> 장거리 경로통행시간 예측결과: 오전 〇형

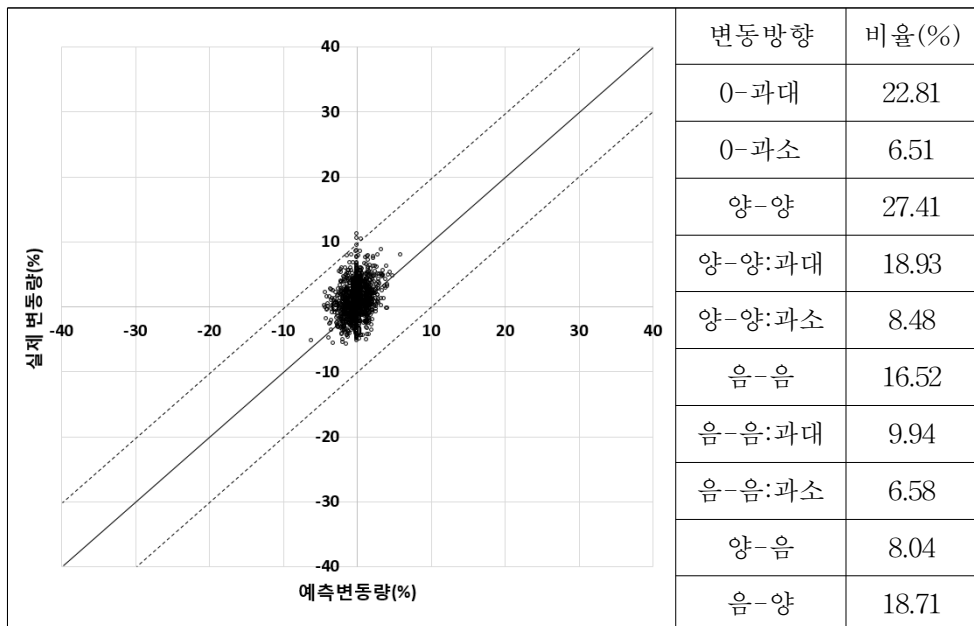


<그림 5-18> 중거리 경로통행시간 예측결과: 오전 〇형

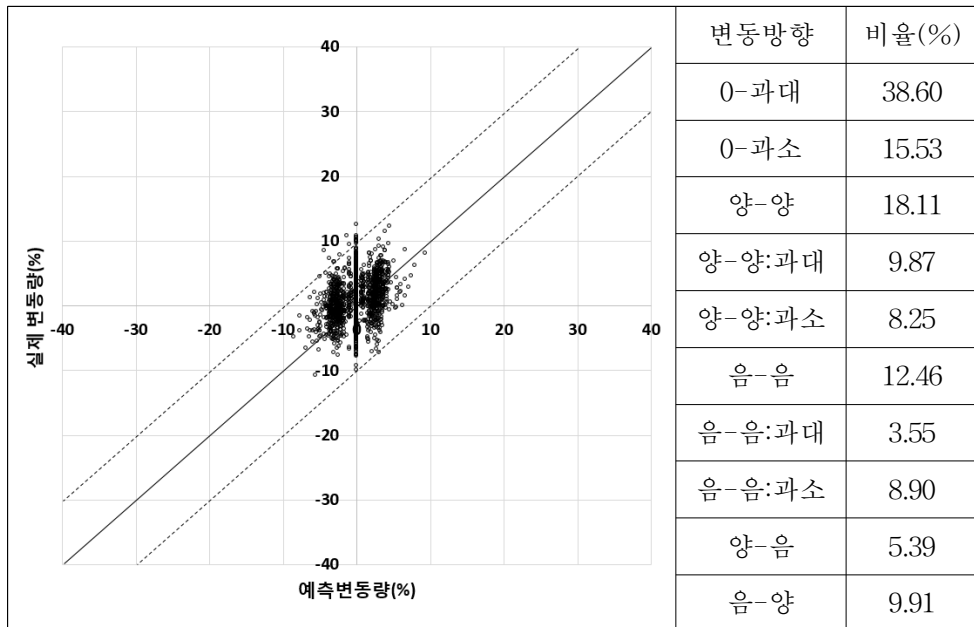


<그림 5-19> 단거리 경로통행시간 예측결과: 오전 ㄱ형

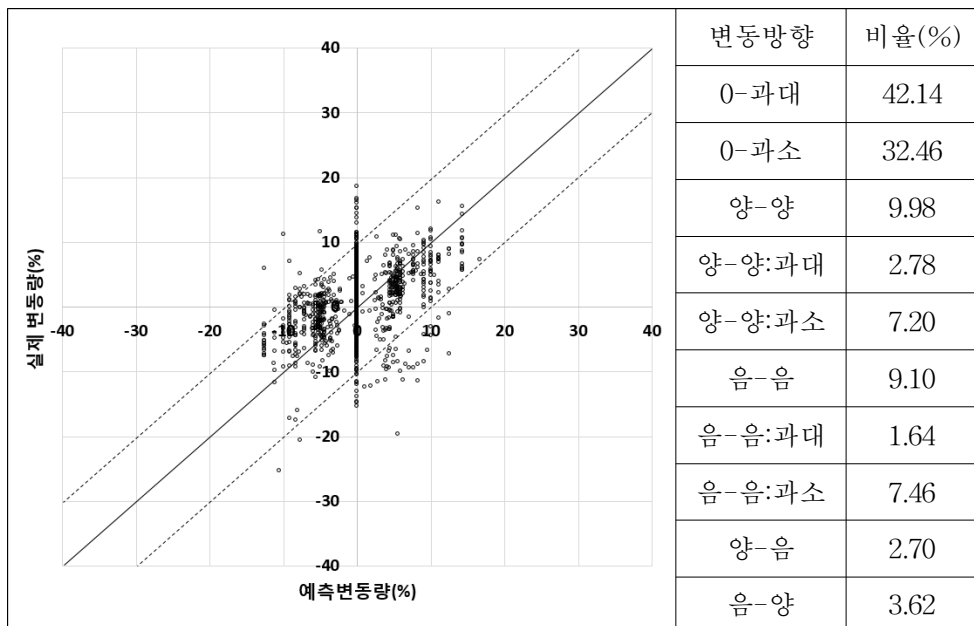
SJ의 특징으로 인하여 [0-과대, 0-과소]의 비율은 장거리, 중거리, 단거리 통행시간 별로 각각 [22.81, 6.51], [38.60, 18.01], [42.14, 32.46]으로 분포하고 있으며, 통행거리가 짧을수록 관측치의 0-과대와 0-과소는 증가하고 있다. 그리고 단거리의 양-양, 음-음의 비율은 유사하게 나타나고 있다. 이는 SJ 혼잡의 시간적 진화과정에서 발생하는 변동 때문이다. 장거리 통행시간의 변동은 원점에 조밀하게 분포하고 있다. 예측변동의 방향성 상실의 아닌 자연적인 변동으로서 실제 통행시간 예측의 관점에서 볼 때 예측력은 우수하다고 판단되며, $\pm 10\%$ 이내의 예측 변동량의 비율은 99.85%로 매우 높게 나타났다. 중거리 통행시간의 관측 변동량은 -4%, 0%, +4%에 분포하고 있으며, 예측 변동량은 순방향 30.57% 역방향 15.3%로 나타나고 있다. 단거리 통행시간의 관측 변동량은 -10~-5%, 0, 5~10%의 3개 영역에 분포함에도 불구하고 예측치의 역변동은 6.32%로 수용수준이며, $\pm 10\%$ 이내의 예측 변동량의 비율은 96.60%로 높게 분석되었다.



<그림 5-20> 장거리 경로통행시간 예측 변동분석: 오전 n형



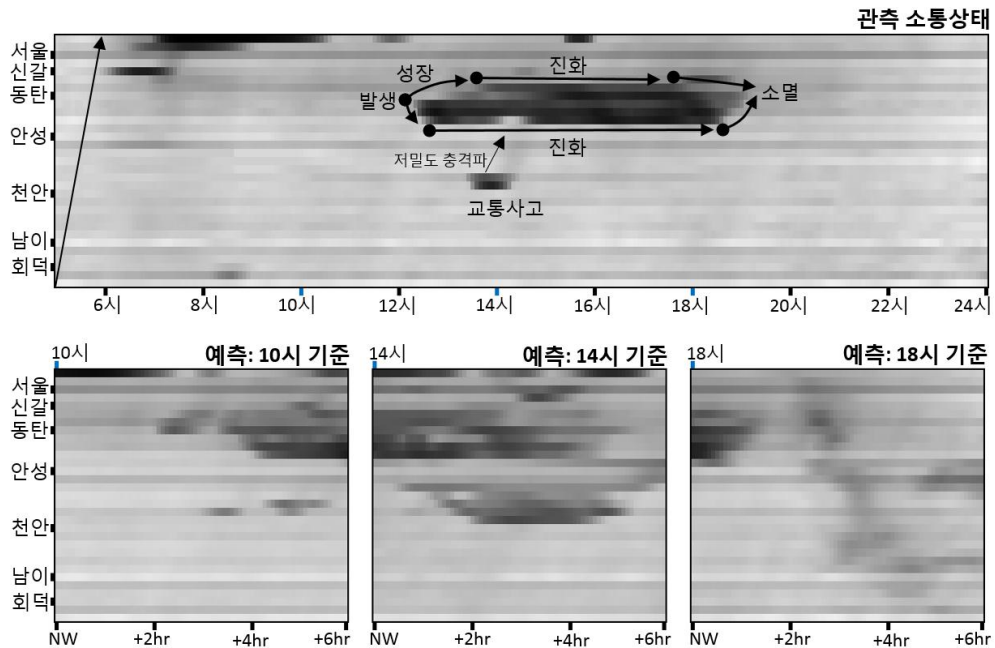
<그림 5-21> 중거리 경로통행시간 예측 변동분석: 오전 n형



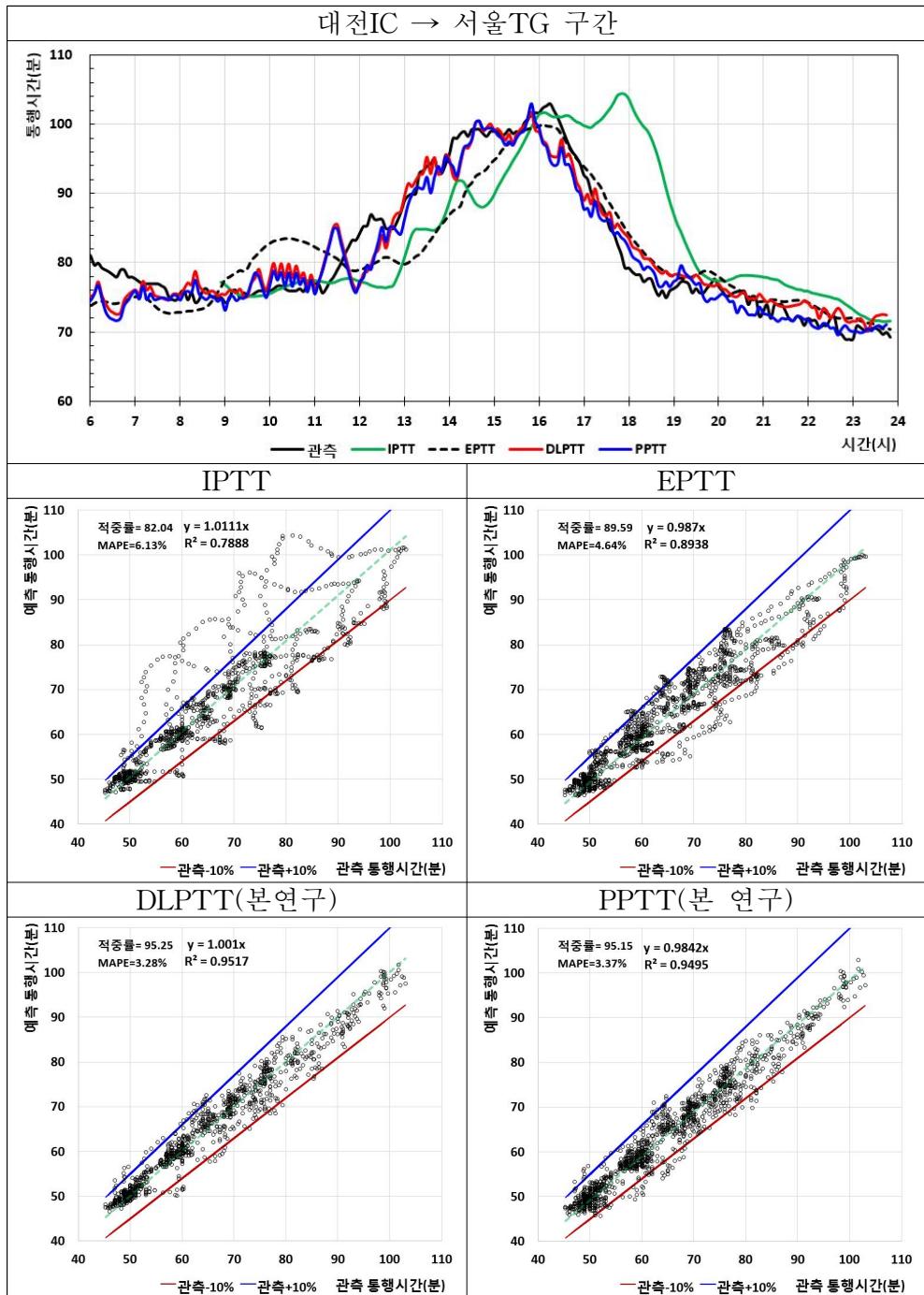
<그림 5-22> 단거리 경로통행시간 예측 변동분석: 오전 n형

4) 적용 결과: 오후 \cap 형I

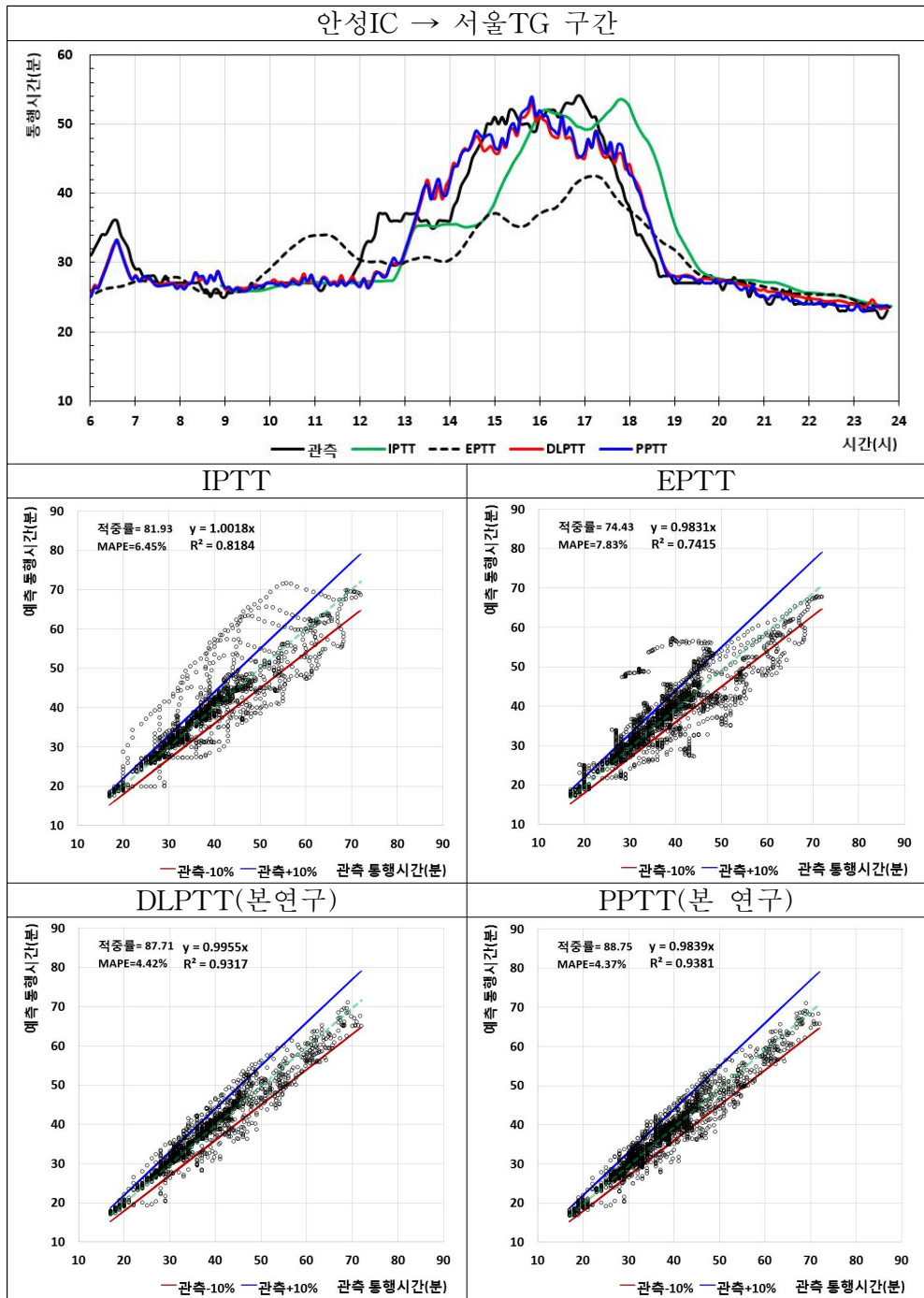
오후 혼잡은 오전에 비하여 반복적 특성이 약하다. 따라서 예측의 관점에서 볼 때 오후 \cap 형 통행시간 형태는 오전 \cap 형에 비하여 불확실성이 높다. 대상 고속도로 구간인 서울-대전의 경우, 오후 혼잡은 서울방향의 차량이 오후에 집중되기 때문에 발생한다. <그림 5-23>에서 예측시점 10시를 기준으로 12시의 혼잡발생과 14시를 기준으로 17시 혼잡소멸을 예측하였다. 그리고 혼잡의 소멸 2시간 전인 18시에 혼잡의 소멸 시간과 위치를 예측하였다. 추가로 14시 천안JC 인근에서 발생한 교통사고로 인해 저밀도 전방충격파의 영향을 받았지만 혼잡의 지속에는 영향을 미치지 못하고 있다. <그림 5-24~26>은 통행거리별 통행시간 예측결과를 보여주고 있으며, IPTT는 상태회복 시간이 길어지면서 낮은 예측력을 보이고 있으며, EPTT는 침두까지의 통행시간 증가를 전혀 설명하지 못하고 있다. 반면, 개발모형은 예측오차 3.37% 이하에서 95.15% 이상의 적중률을 보이고 있다.



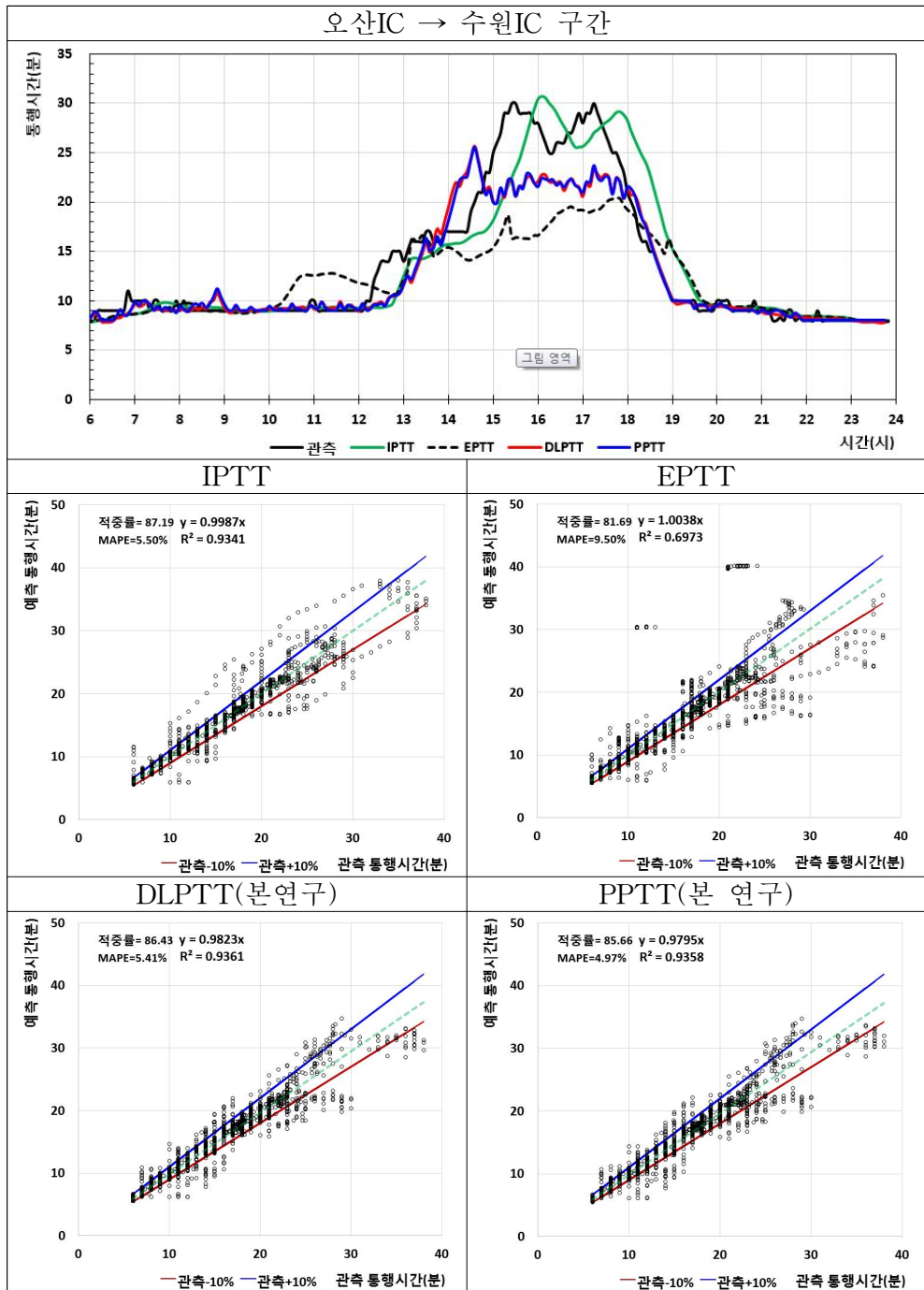
<그림 5-23> 소통상태 예측결과: 오후 \cap 형I



<그림 5-24> 장거리 경로통행시간 예측결과: 오후 n형

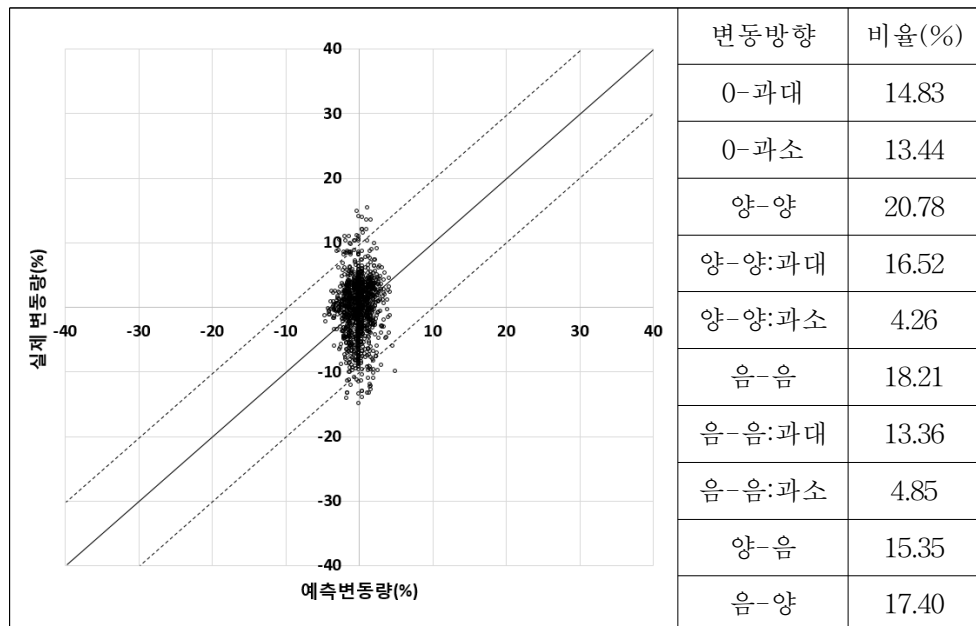


<그림 5-25> 중거리 경로통행시간 예측결과: 오후 ㄱ형

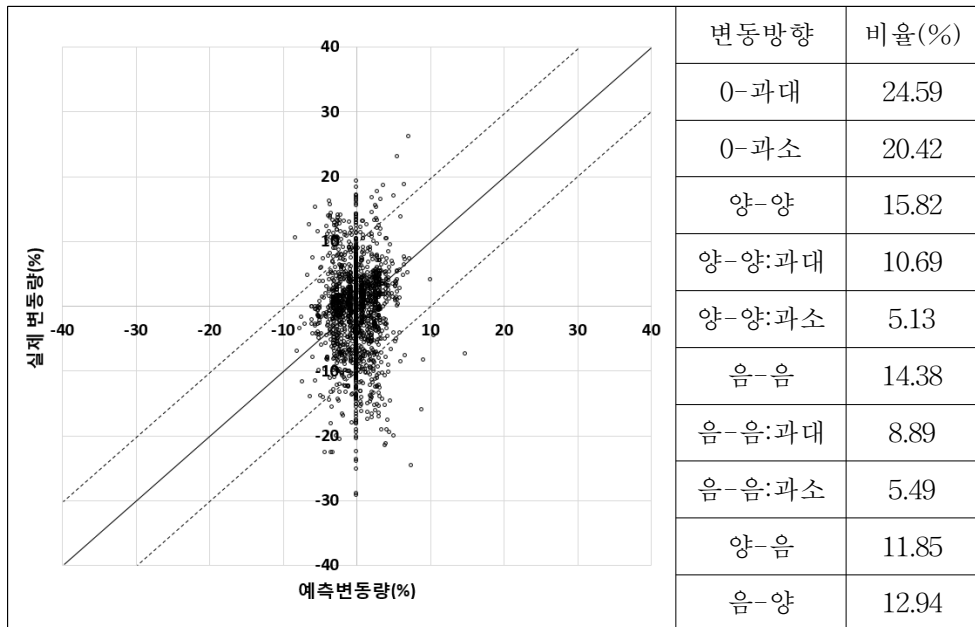


<그림 5-26> 단거리 경로통행시간 예측결과: 오후 n형

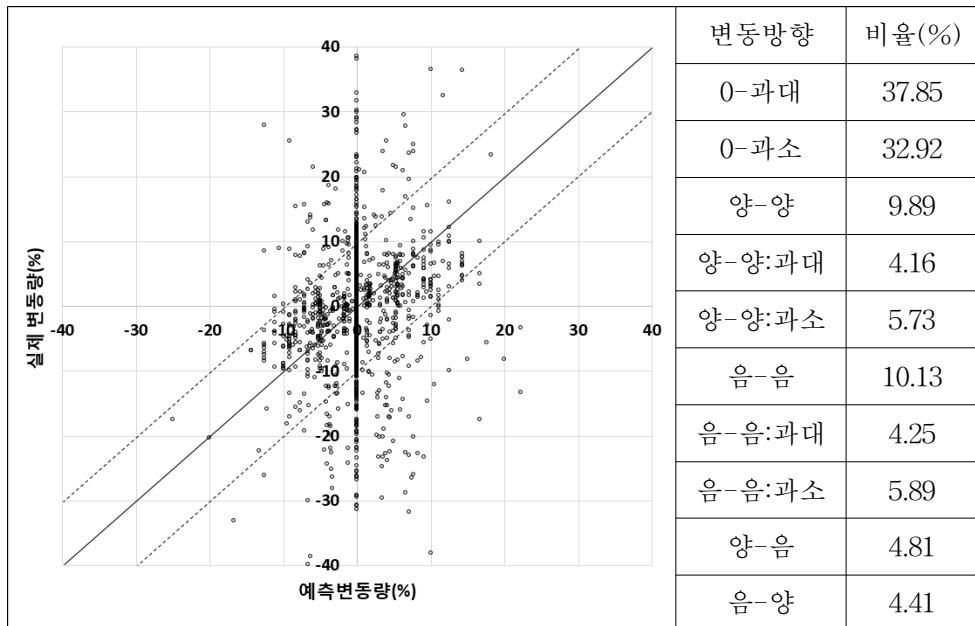
SJ의 특징으로 인하여 [0-과대, 0-과소]의 비율은 통행거리가 감소함에 따라 증가하고 있다. 단거리 통행의 관측 변동량이 0인 구성비는 70.77%에 달하고 있다. 통행거리가 증가함에 따라 관측 변동량의 폭은 $\pm 3\%$ 에서 $\pm 10\%$ 까지 증가하고 있으며, 예측 변동량의 폭은 $\pm 10\%$ 에서 $\pm 30\%$ 까지 증가하고 있다. 따라서 장·중·단거리 통행에 대한 $\pm 10\%$ 이내 예측 변동의 구성비(%)는 각각 95.23, 88.47, 86.30으로 통행거리가 감소함에 따라 감소하고 있다. 일반적으로 예측 변동량 축을 따라 세로로 변동점이 분포한 경우의 예측력은 낮지만 원점에 집중적으로 변동점이 형성된 세로형 형태의 변동분포 경우 예측력은 높다고 말할 수 있다. IPTT와 EPTT가 상태의 증가와 감소의 과정에서 예측력을 거의 상실한 것에 비하면 개발 모형의 예측력은 높다고 판단된다. 변동 분포의 특성분석 결과를 기초로 일정 공간을 지속적으로 점유하는 SJ의 시간적 진화행태는 다양하게 나타남을 간접적으로 확인 할 수 있다. 그리고 일정 공간에 갇힌 시간적 진화는 통행시간의 침두 전·후에서 시계열의 변동으로 나타난다.



<그림 5-27> 장거리 경로통행시간 예측 변동분석: 오후 n형



<그림 5-28> 중거리 경로통행시간 예측 변동분석: 오후 n형I

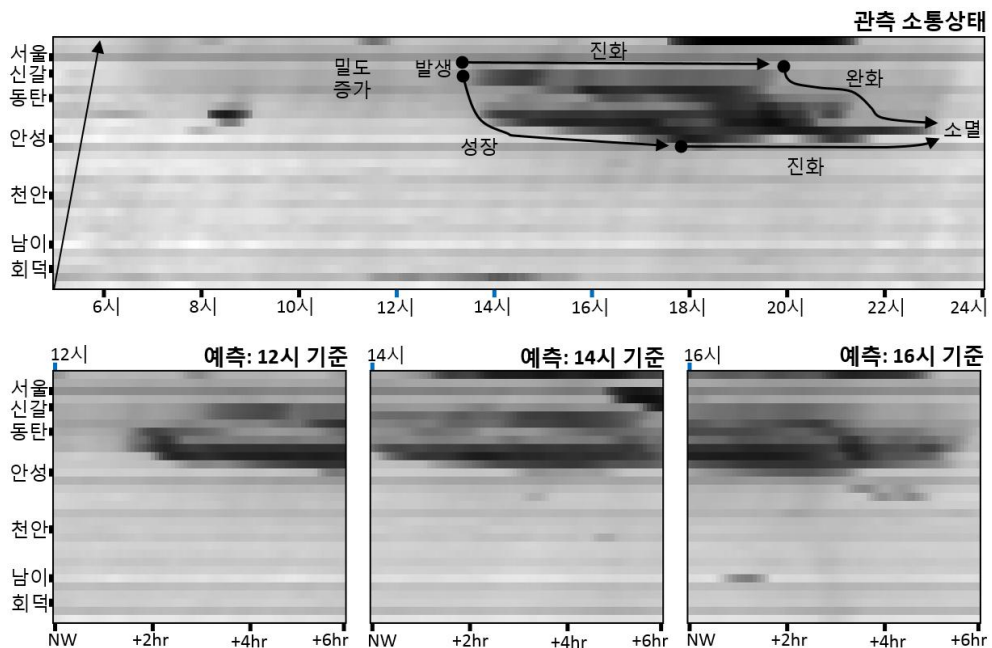


<그림 5-29> 단거리 경로통행시간 예측 변동분석: 오후 n형I

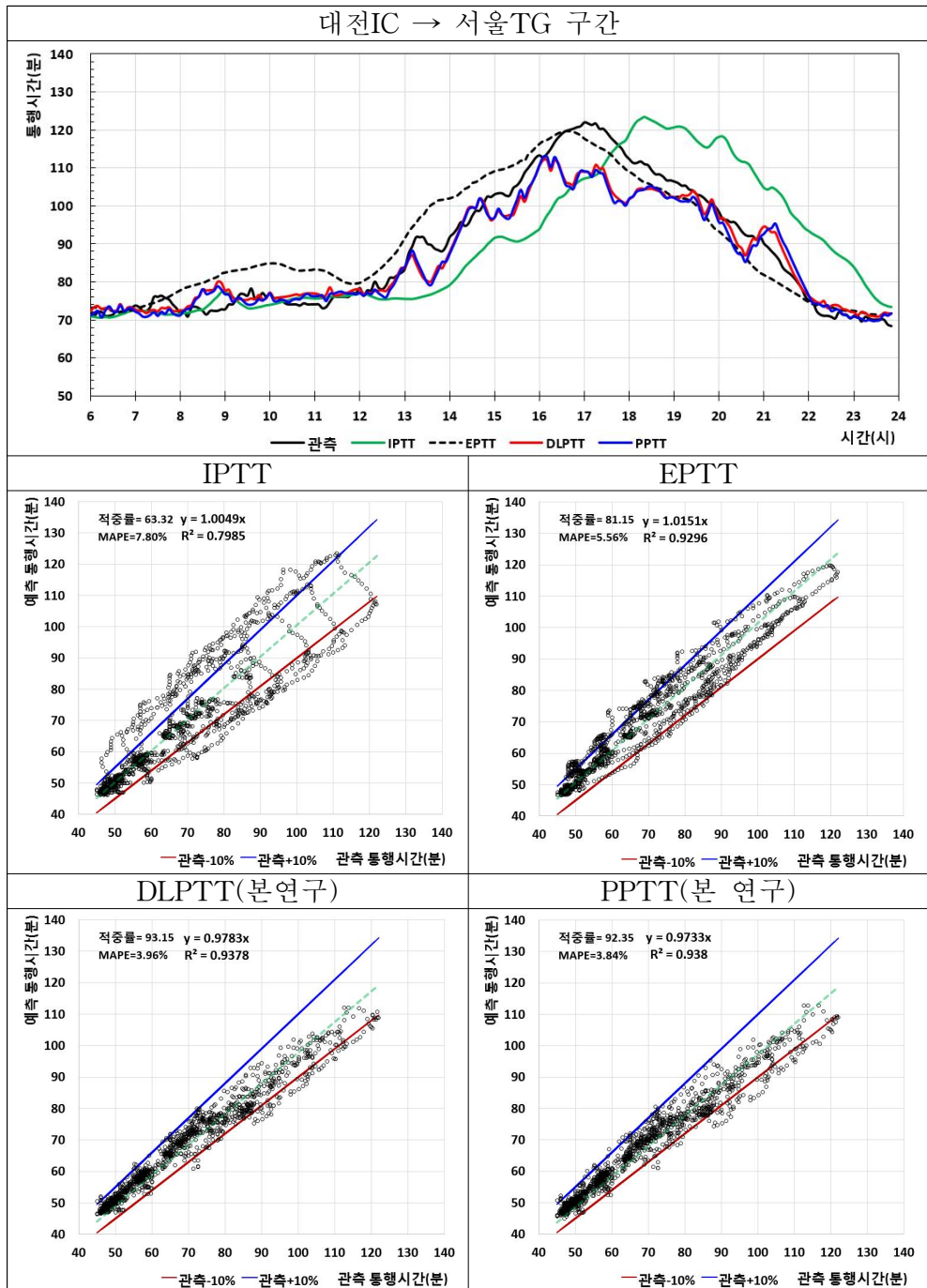
5) 적용 결과: 오후 Ⅱ형Ⅱ

오후 Ⅱ형Ⅱ의 통행시간 형태는 <그림 5-30>과 같이 중밀도 규모의 혼잡이 충분한 도로공간을 점유하면서 지속적으로 발생-성장-진화-완화-소멸하는 과정에서 나타난다. 신갈JC 부근에서 발생한 혼잡은 안성JC 까지 확장된 후 9시간 동안 지속된 후 소멸하고 있다. 예측시점 12시 기준으로 2시간 후 중밀도 혼잡발생을 예측하였으며, 14시 시점에서 6시간 혼잡의 지속과 더불어 1시간 후 판교JC 부근의 혼잡을 예측하였다. 그리고 예측시점 16시를 기준으로 6시간 후 혼잡의 완전소멸을 예측하였다.

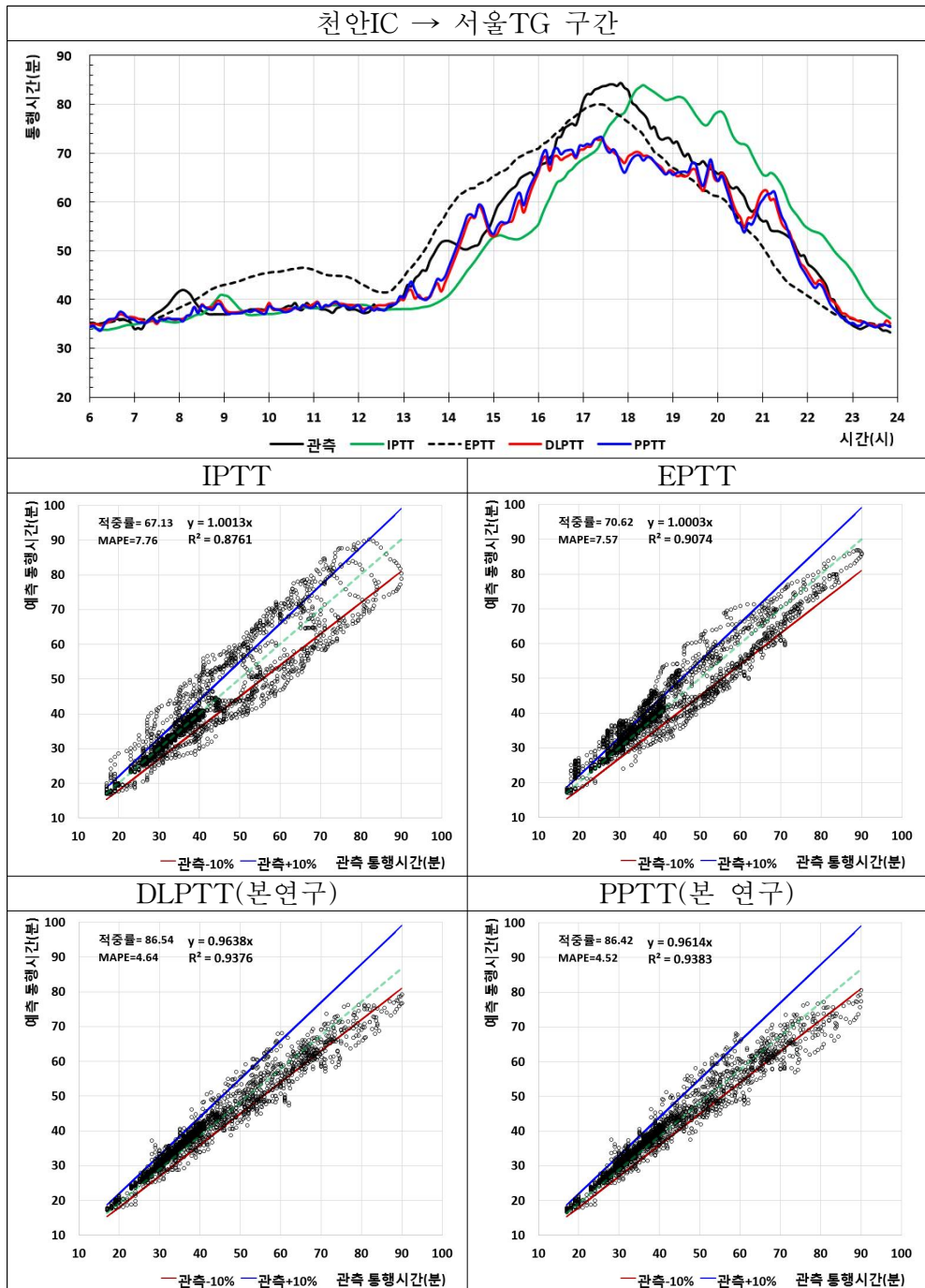
<그림 5-31>에서 4개의 모형중 2개 개발모형의 예측오차와 적중률이 가장 우수하게 나타났으며, IPTT의 경우 완만한 형태의 통행시간 증가와 감소에도 불구하고 상태회복 시간은 길어지면서 예측력은 현저히 감소하고 있으며, 사실상 예측실패를 보이고 있다. 반면, 개발모형의 경우 예측오차는 3.96% 이하로 나타나고 있으며 적중률은 92.35% 이상으로서 비교모형과 큰 차이를 보이고 있다.



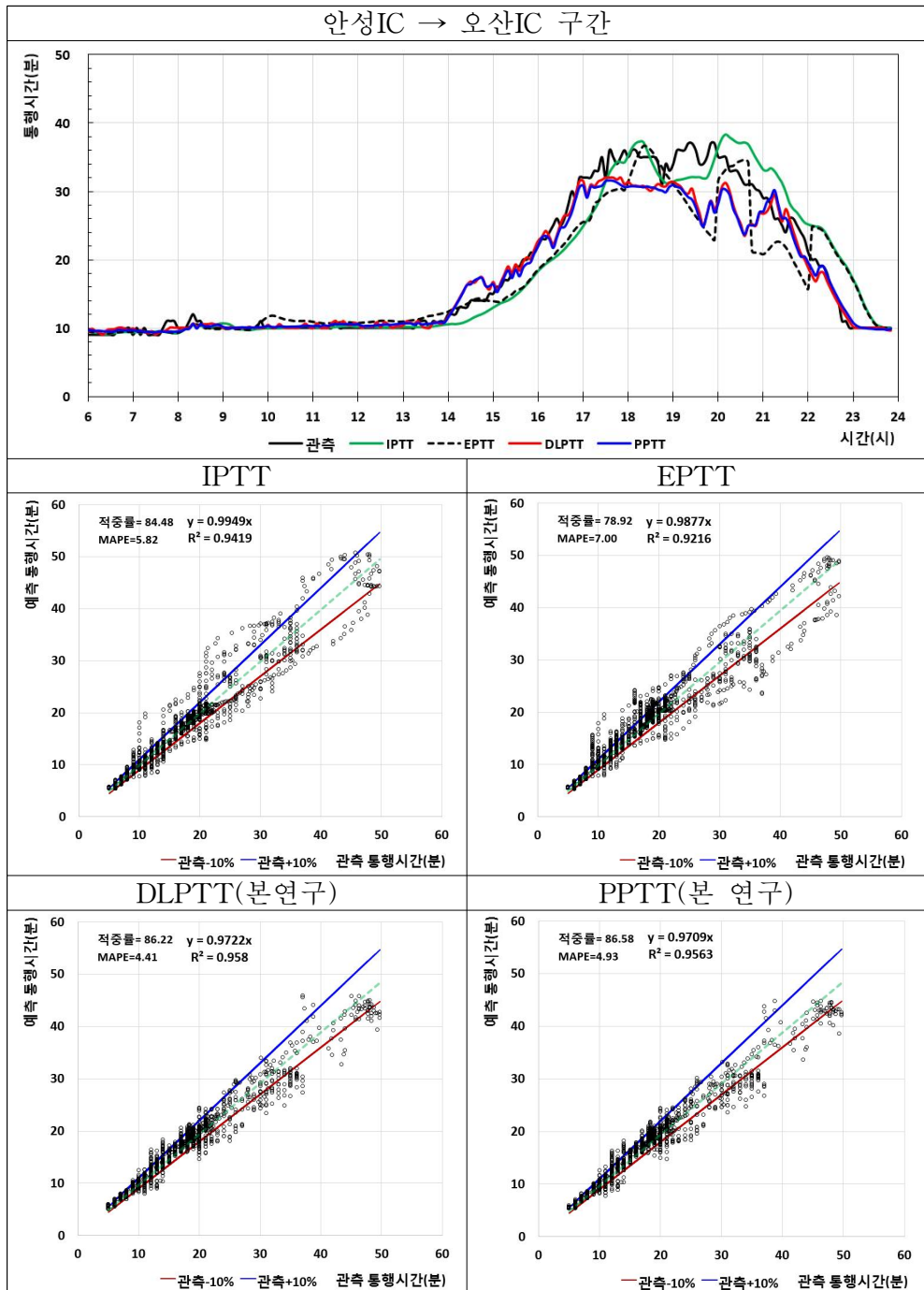
<그림 5-30> 소통상태 예측결과: 오후 Ⅱ형Ⅱ



<그림 5-31> 장거리 경로통행시간 예측결과: 오후 n형II

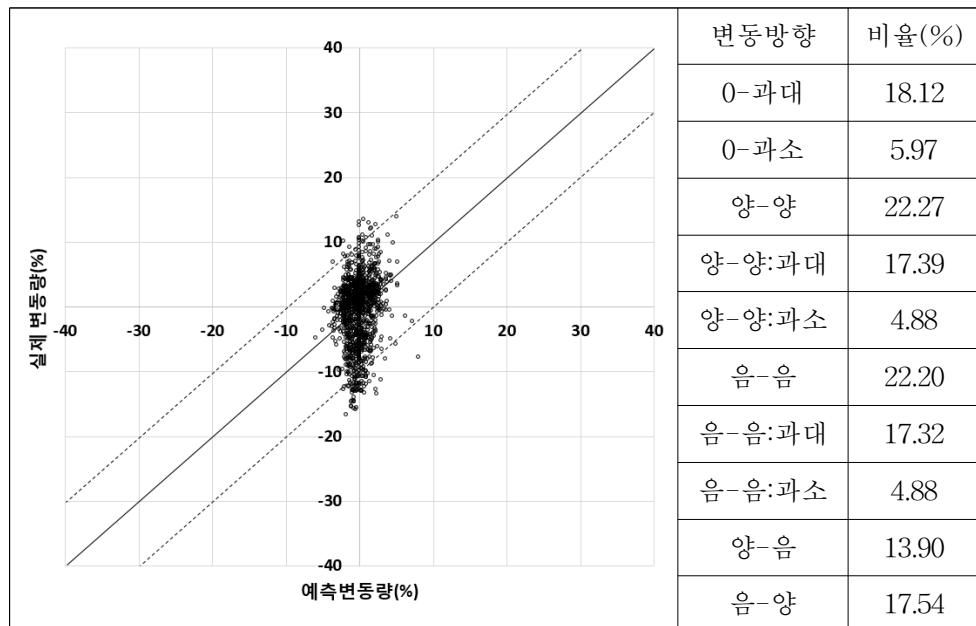


<그림 5-32> 중거리 경로통행시간 예측결과: 오후 n형II

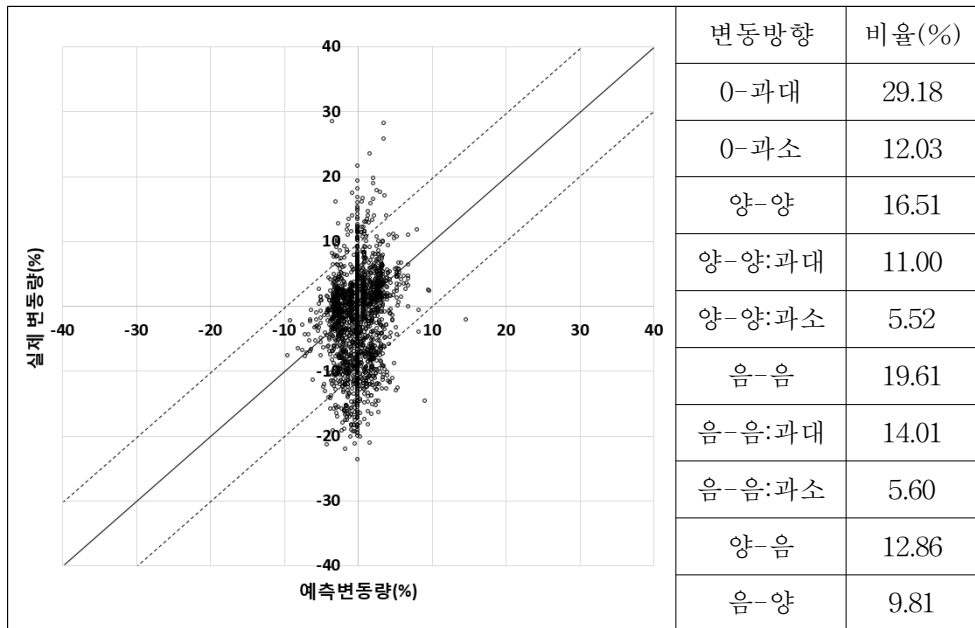


<그림 5-33> 단거리 경로통행시간 예측결과: 오후 ∩형II

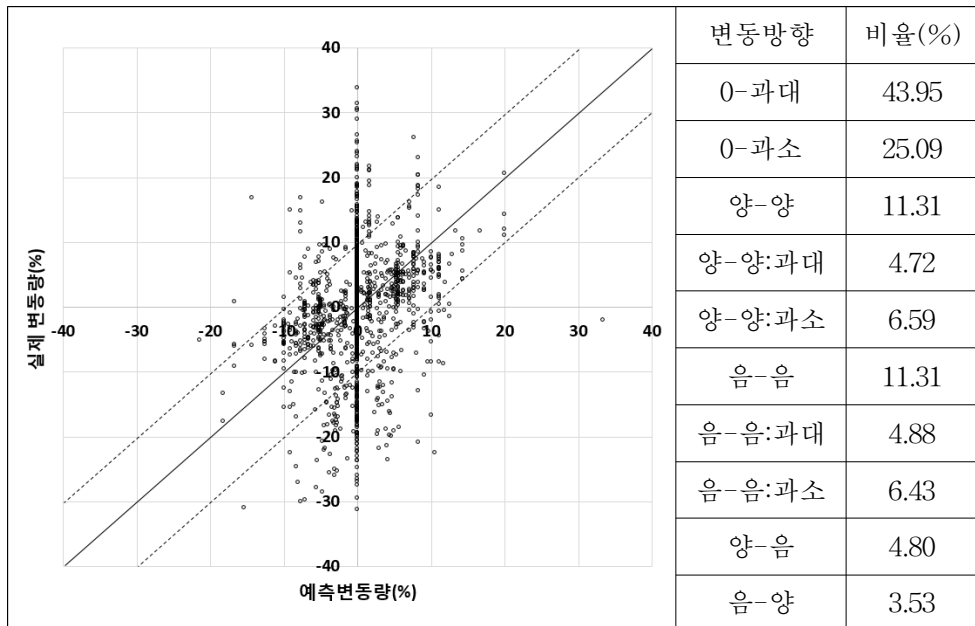
변동점의 형태는 통행거리가 증가함에 따라 퍼진 세로형에서 좁혀진 세로형으로 나타나고 있다. 이는 장거리 통행시간의 경우 절대적인 통행시간의 양이 크기 때문에 실제 변동량(%)이 낮게 나타나기 때문이다. 단 · 중 · 장거리 통행시간의 $\pm 10\%$ 이내 예측 변동량의 구성비(%)는 86.34, 86.30, 93.16으로 나타나고 있다. 이는 실제 변동량에 대한 순방향의 과대/과소 추정문제와 더불어 역변동이 증가할 수 있음을 의미한다. 역변동의 구성비(%)는 단 · 중 · 장거리 통행시간에 대하여 각각 8.33, 22.67, 31.44로서 장거리의 방향성 상실이 높게 나타났다. 그러나 단거리 통행시간의 절대적인 양의 작기 때문에 변동의 방향성을 설명함에도 불구하고 예측오차는 중 · 장거리 통행시간보다 높게 나타났다. 중 · 장거리 통행시간의 순방향 변동은 과대가 과소보다 5~10% 이상 크게 나타난 반면, 단거리 통행시간의 순방향 변동은 2%의 차이를 보이고 있지 않다. 이는 통행거리가 증가함에 따라 변동의 방향에 대한 모형의 설명력이 저하됨을 의미하며, <그림 5-31~33> 또한 동일한 결과를 보이고 있다.



<그림 5-34> 장거리 경로통행시간 예측 변동분석: 오후 Ⅱ형II



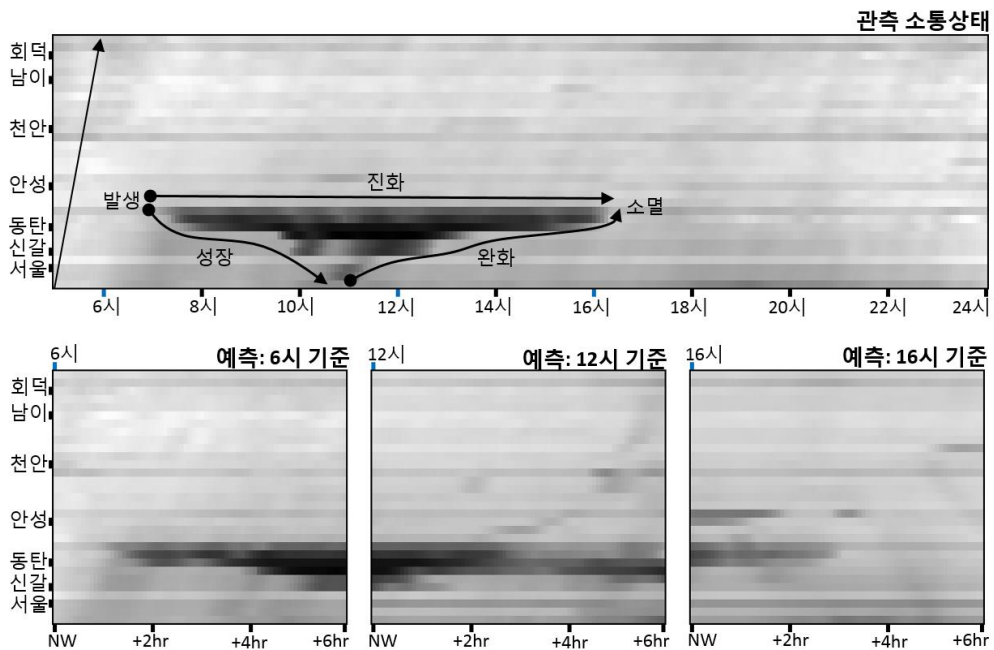
<그림 5-35> 중거리 경로통행시간 예측 변동분석: 오후 Ⅱ형II



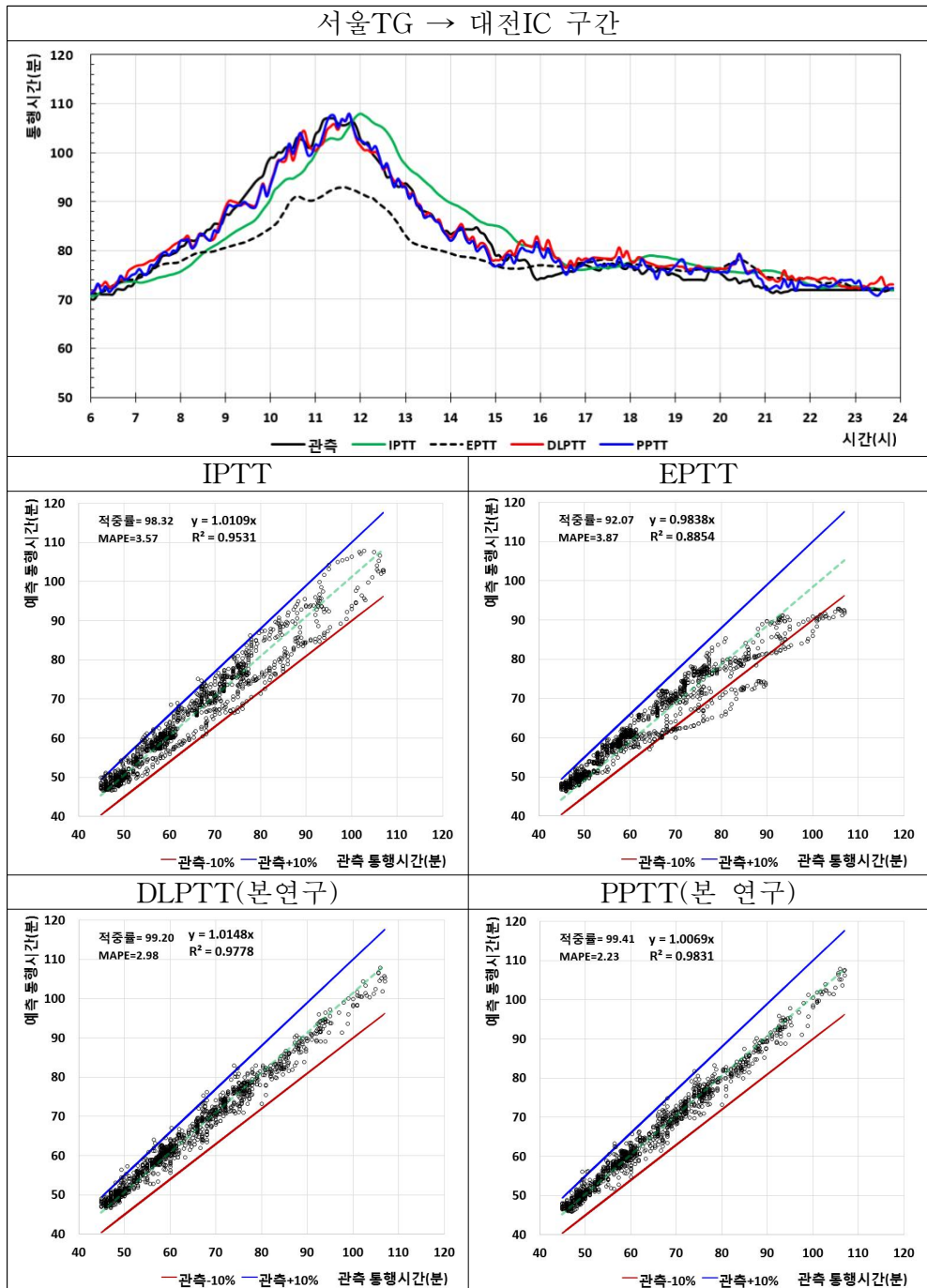
<그림 5-36> 단거리 경로통행시간 예측 변동분석: 오후 Ⅱ형II

6) 적용 결과: 오전 \wedge 형

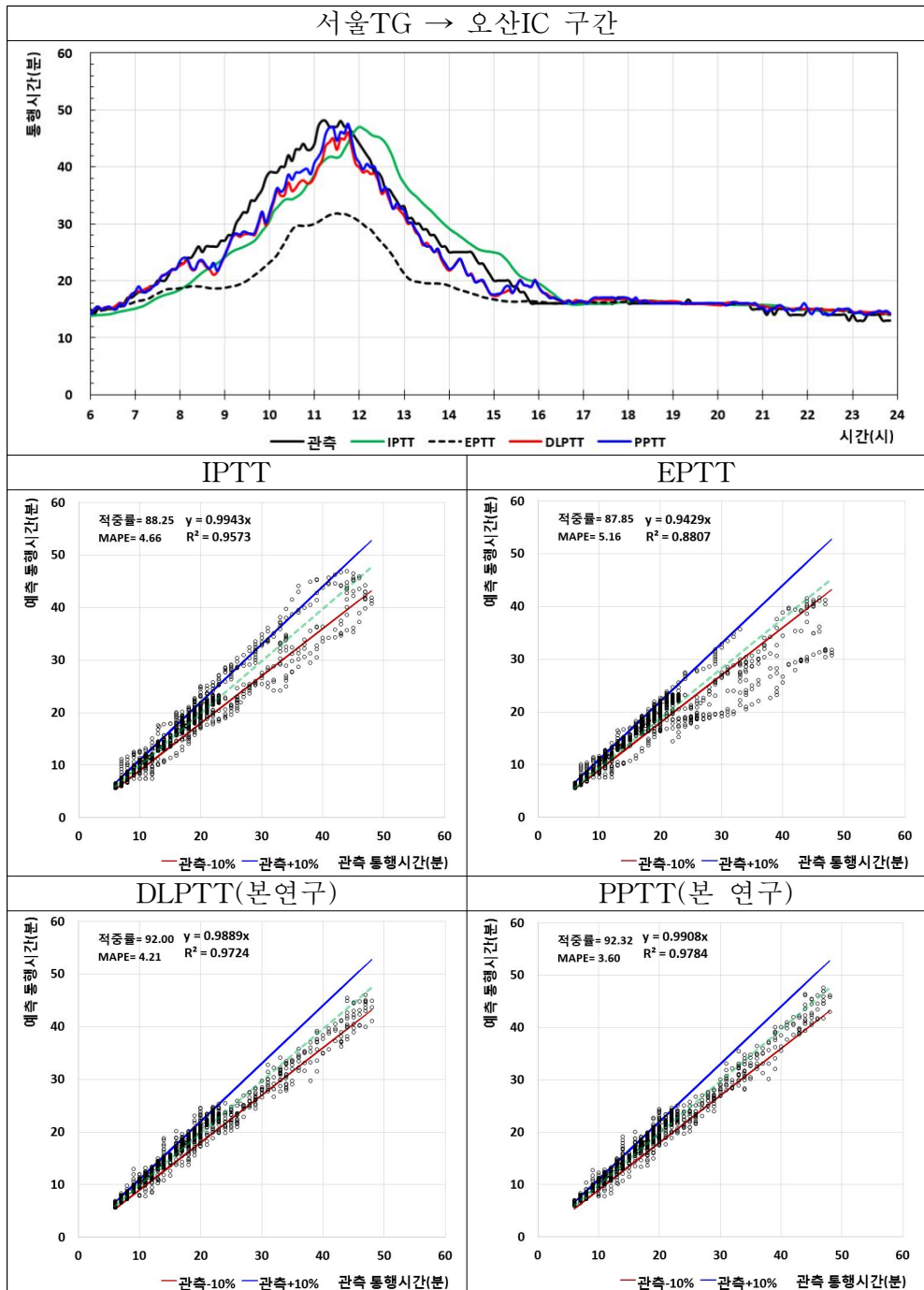
\wedge 형 통행시간 형태는 통행시간이 증가한 후 짧은 침두상태를 유지한 후 다시 감소하는 형태이다. 따라서 \wedge 형 통행시간 형태는 <그림 5-37>과 같이 혼잡이 공간적으로 성장한 후 시간적 진화의 과정없이 완화상태를 거쳐 소멸하게 됨으로 혼잡의 형태는 ∇ 의 형태를 보인다. 예측시점 6시를 기준으로 장래 6시간 동안 혼잡의 발생과 지속을 예측하고 있으며, 12시를 기준으로 4시간후 혼잡의 완화를 예측하고 있다. 그리고 16시 이후 혼잡이 발생하지 않음을 혼잡의 발생-성장-완화-소멸단계에 걸쳐 정확히 예측하고 있다. 따라서 <그림 5-36>에서 DLPTT와 PPTT의 예측오차는 2.98%, 2.23%으로 매우 낮게 나타나고 있으며 적중률은 99.20% 이상으로 높게 나타나고 있다. 반면, 혼잡시 EPTT의 경우 방향성은 설명하지만 변동량을 설명하지 못 하기 때문에 통행시간 예측오차는 실제 통행시간이 증가함에 따라 증가하고 있다. 이는 반복적 특성을 보이는 혼잡이더라고 혼잡의 강도와 내부 진화의 행태가 다르게 나타나며, 개발 모형은 이러한 내재적 특성을 인식한다는 간접적 증거이다.



<그림 5-37> 소통상태 예측결과: 오전 \wedge 형

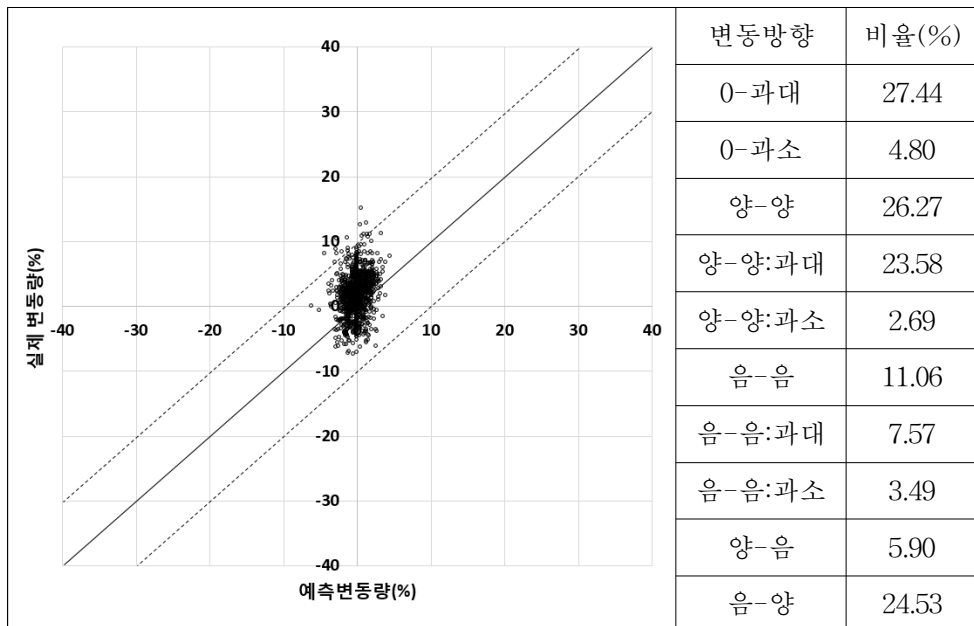


<그림 5-38> 장거리 경로통행시간 예측결과: 오전 ^형

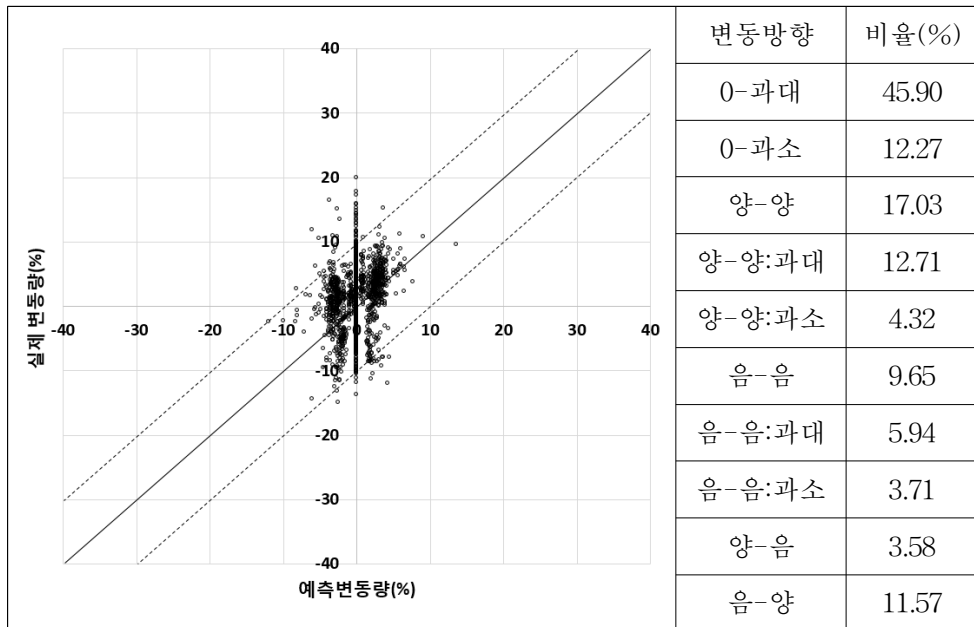


<그림 5-40> 단거리 경로통행시간 예측결과: 오전 ^형

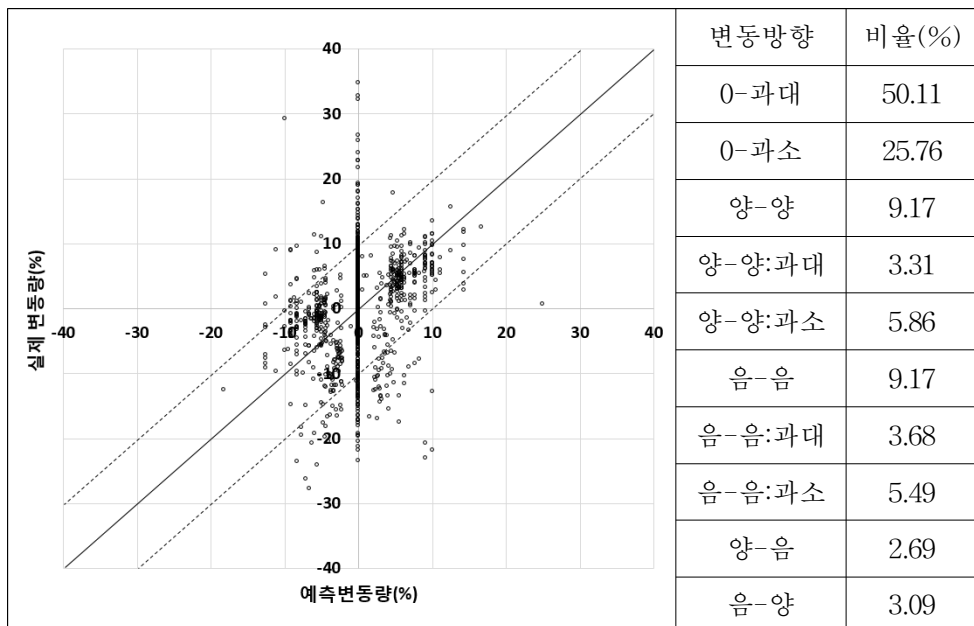
단·중·장거리 통행시간의 $\pm 10\%$ 이내 예측 변동량의 구성비(%)는 91.85, 97.03, 99.12로 나타났다. 장거리 통행시간의 변동 형태는 0-과대 27.44%, 양-양 과대 26.27%, 음-양의 역변동 24.53%로서 예측 변동량의 양방향 구성비는 75.55%로 나타났다. 그러나 변동량은 $-3 \sim +5\%$ 이내에 분포하고 있어 예측력은 우수하다고 판단된다. 중거리 통행시간의 변동 형태는 실제 변동량 -3% , 0% , $+3\%$ 에서 3개의 세로형 변동을 보이고 있다. 예측 변동량은 양의 과대방향으로 편중되어 있으며, 0-과대 45.90%, 양-양 과대 12.71%, 음-양 11.57%로 높게 나타났다. 그러나 예측 변동량은 대부분 $-9 \sim +9\%$ 이내로 나타나 수용가능한 수준이라 판단된다. 단거리 통행시간의 실제 변동량은 -5% , 0% , $+5\%$ 에서 세로형 변동을 각각 보이고 있다. 0 변동은 75.77% 높게 나타났으며, 이는 단거리 통행시간의 혼잡이 매우 낮았음을 의미한다. 그리고 양-양과 음-음의 순방향 변동은 각각 9.17%로 나타났으며 완전 예측선을 따라 분포하고 있다.



<그림 5-41> 장거리 경로통행시간 예측 변동분석: 오전 ^형



<그림 5-42> 중거리 경로통행시간 예측 변동분석: 오전 ^형

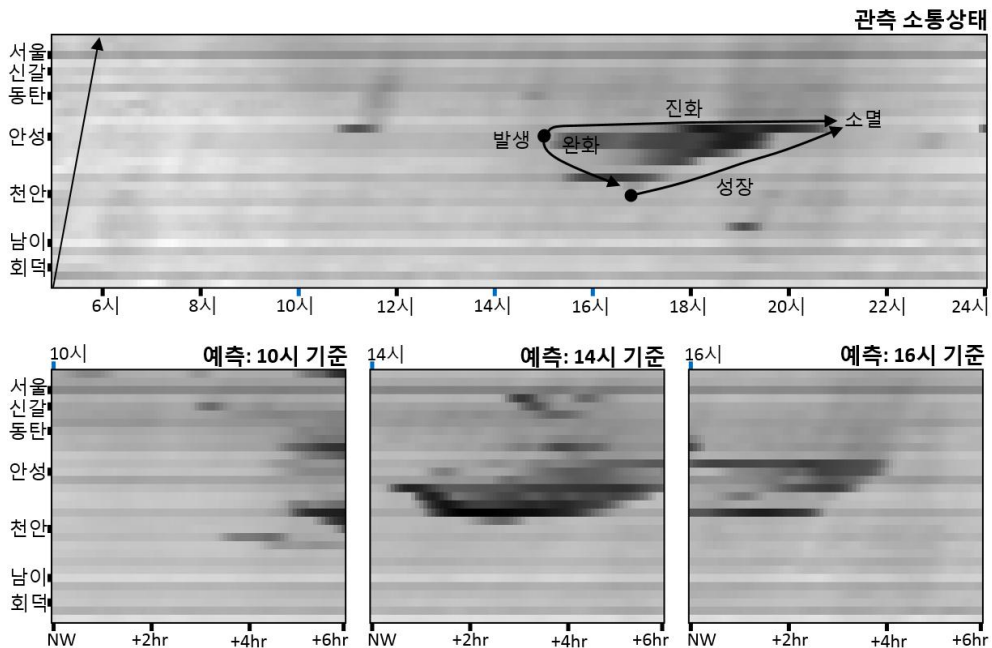


<그림 5-43> 단거리 경로통행시간 예측 변동분석: 오전 ^형

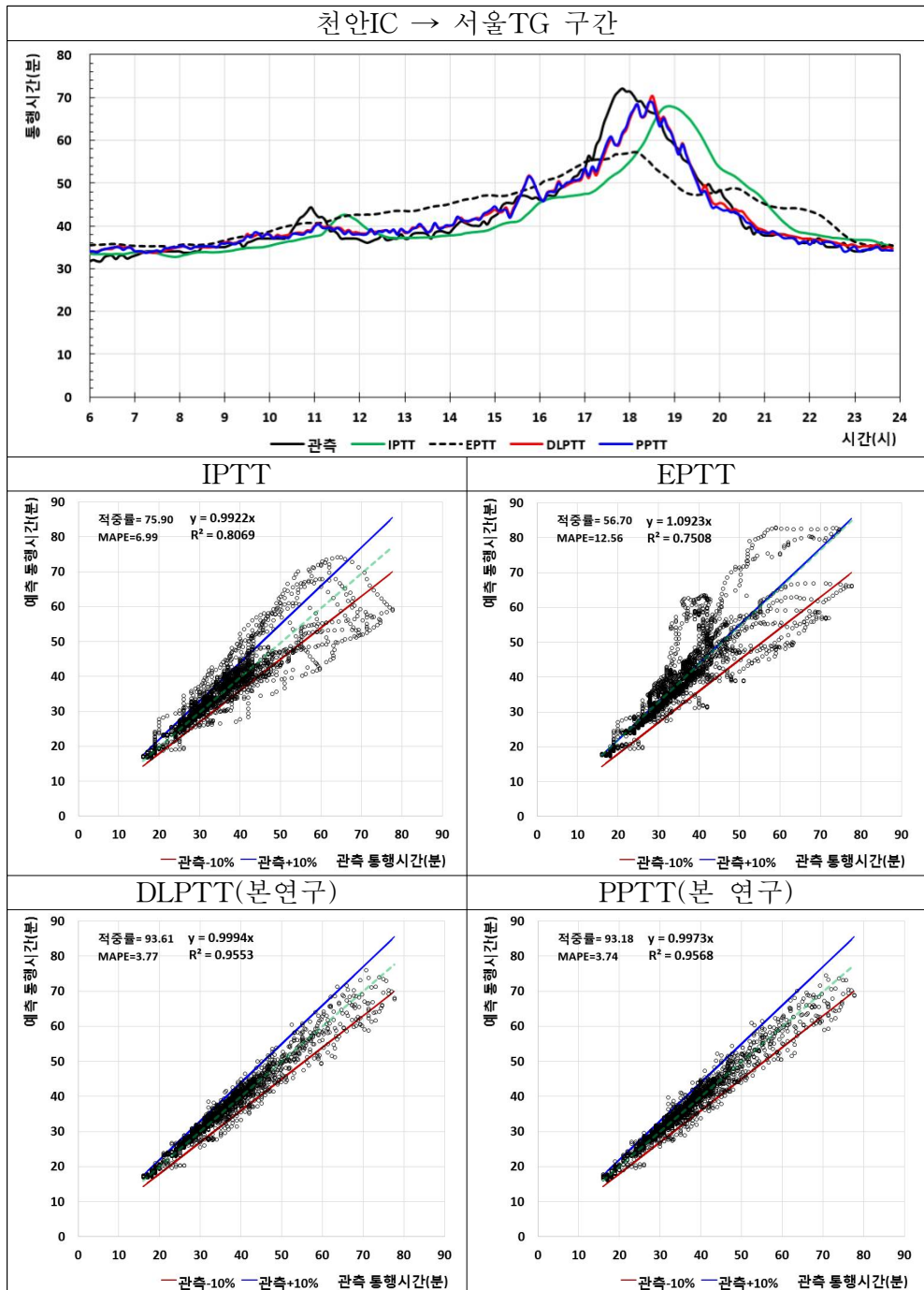
7) 적용 결과: 오후 ㄱ형

ㄱ형 통행시간은 일반적으로 ㄴ형태의 혼잡에서 나타나며, ㄴ형 통행시간에 비하여 혼잡의 생애주기는 길지 않은 편이다. 그리고 통행시간의 상승과 하강 곡선의 기울기는 ㄴ형 비하여 가파르게 나타난다. <그림 5-44>에서 예측모형은 10시를 기준으로 5시간 이후의 혼잡을 예측하고 있으며, 14시에는 약 1시간이후 혼잡발생을 예측하고 있다. 혼잡의 발생 이후, 16시를 기준으로 4시간후 혼잡소멸을 어느정도 예측하고 있다.

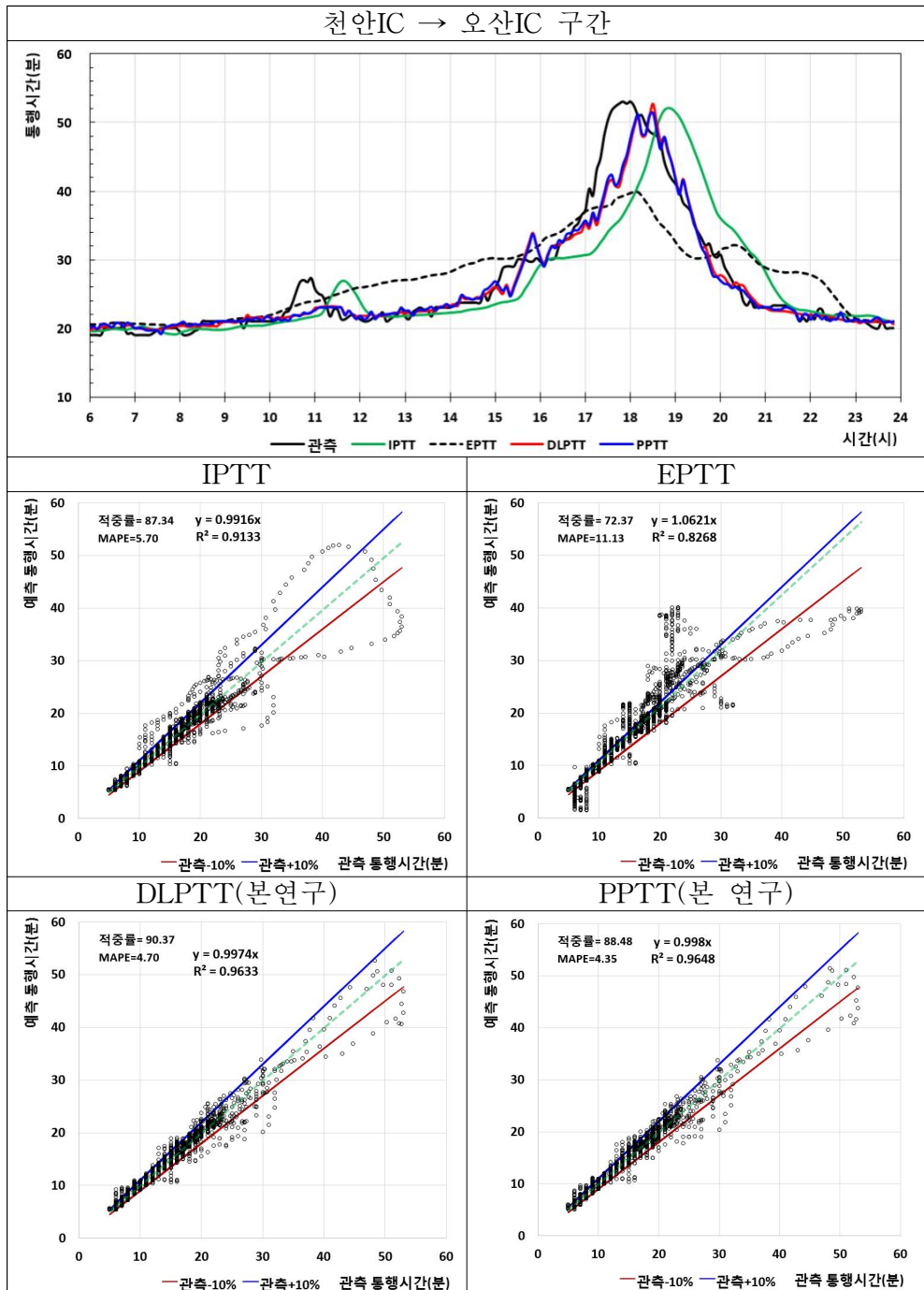
<그림 5-45>의 통행속도 예측결과에서 개발모형의 예측오차는 3.72% 이하로 나타나고 있으며, 적중률은 95%을 넘고 있다. 반면 IPTT와 EPTT의 적중률은 각각 76.49%, 43.01%로 나타나고 있다. IPTT이 예측력이 낮게 나타난 이유는 ㄱ형태에서 발생하는 급격한 상태변동을 설명하지 못 하기 때문이다. EPTT의 예측오차가 높게 나타나는 원인은 이력 평균의 구축시 ㄱ은 ㄴ형태로 평활화되고 ㄱ형태 혼잡은 시공간적으로 국부적인 경우가 많기 때문에 EPTT는 이를 반영하기 어렵다.



<그림 5-44> 소통상태 예측결과: 오후 ㄱ형

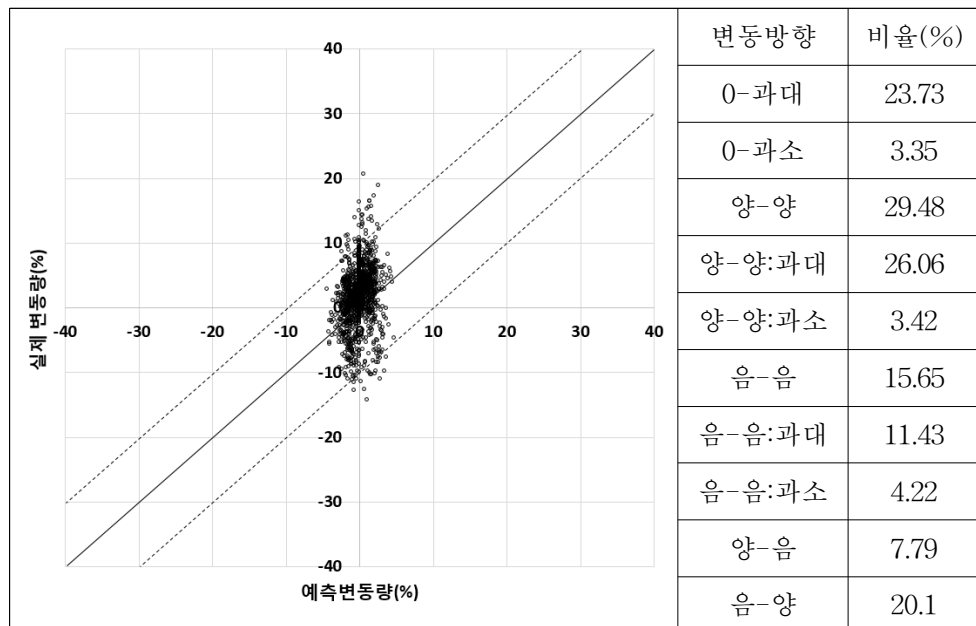


<그림 5-46> 중거리 경로통행시간 예측결과: 오후 ^형

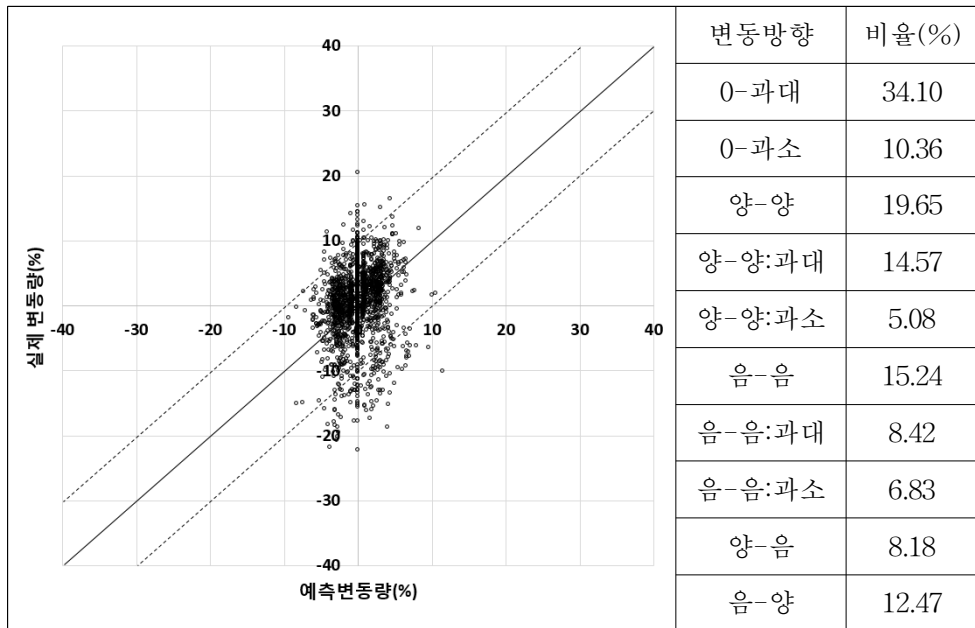


<그림 5-47> 단거리 경로통행시간 예측결과: 오후 ^형

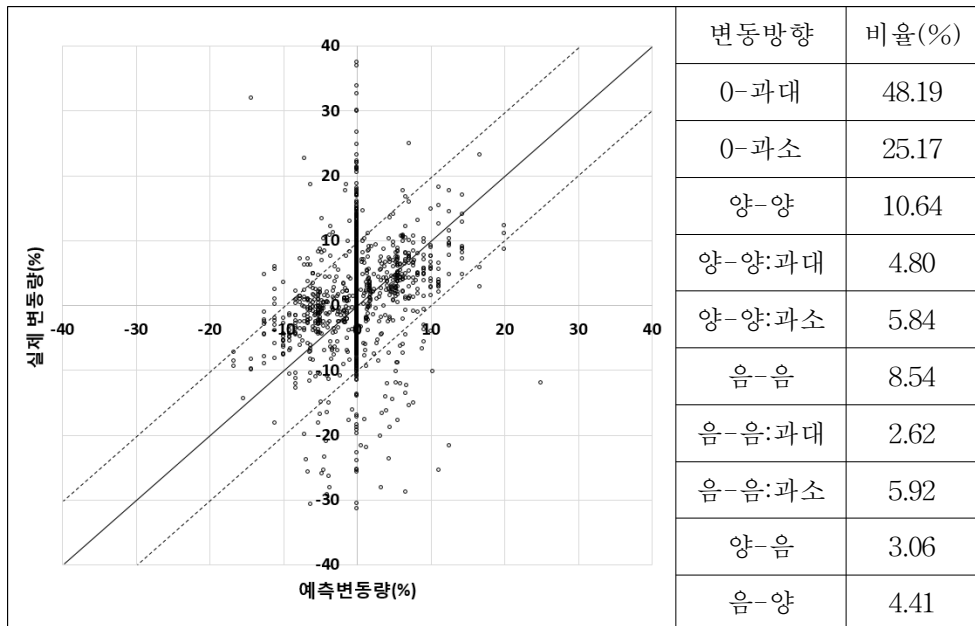
장거리 통행시간의 변동 형태는 0-과대 23.73%, 0-과소 3.35%로 각각 나타나 혼잡을 경험하지 않은 장거리 통행은 27.08%로 나타났으며, 양-양 29.48% 음-음 15.65%로 나타나 실제 변동비와 예측 변동비의 방향성이 일치한 비율은 45.13%로서 역변동인 양-음 7.79% 음-양 20.1% 보다 높은 구성비를 차지하였다. $\pm 10\%$ 이내 예측 변동량은 95.78%로 나타나 우수한 예측력을 보였다고 판단된다. 중거리 통행시간의 변동 형태는 실제 변동량 -3%, 0%, +3%에서 3개의 세로형 변동을 보이고 있으며, 0-과대와 0-과소의 합은 44.46%로서 혼잡을 경험하지 않은 통행은 장거리 통행보다 높게 나타났다. 실제 변동량이 +인 경우 예측 변동량의 26.06%가 +방향으로 집중되어 있으며, 실제 변동량이 -인 경우 예측 변동량의 15.24%가 -방향으로 집중되어 있다. 따라서 변동의 형태는 타원형이지만 완전 예측선에 집중된 형태를 보이고 있으며 $\pm 10\%$ 이내 예측 변동량은 93.69%로 나타났다. 단거리 통행시간은 실제 변동량이 0인 구성비가 73.36%로 매우 높게 나타났으며 역변동은 7.47%로 나타나 완전 예측선에 퍼진 형태로 분포하고 있다.



<그림 5-48> 장거리 경로통행시간 예측 변동분석: 오후 ^형



<그림 5-49> 중거리 경로통행시간 예측 변동분석: 오후 ^형

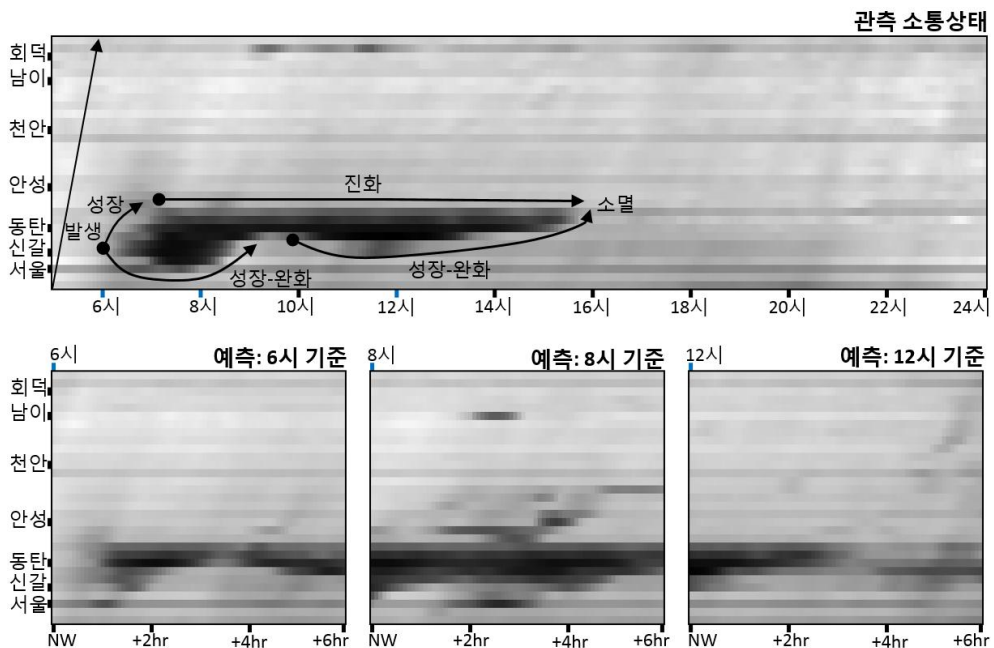


<그림 5-50> 단거리 경로통행시간 예측 변동분석: 오후 ^형

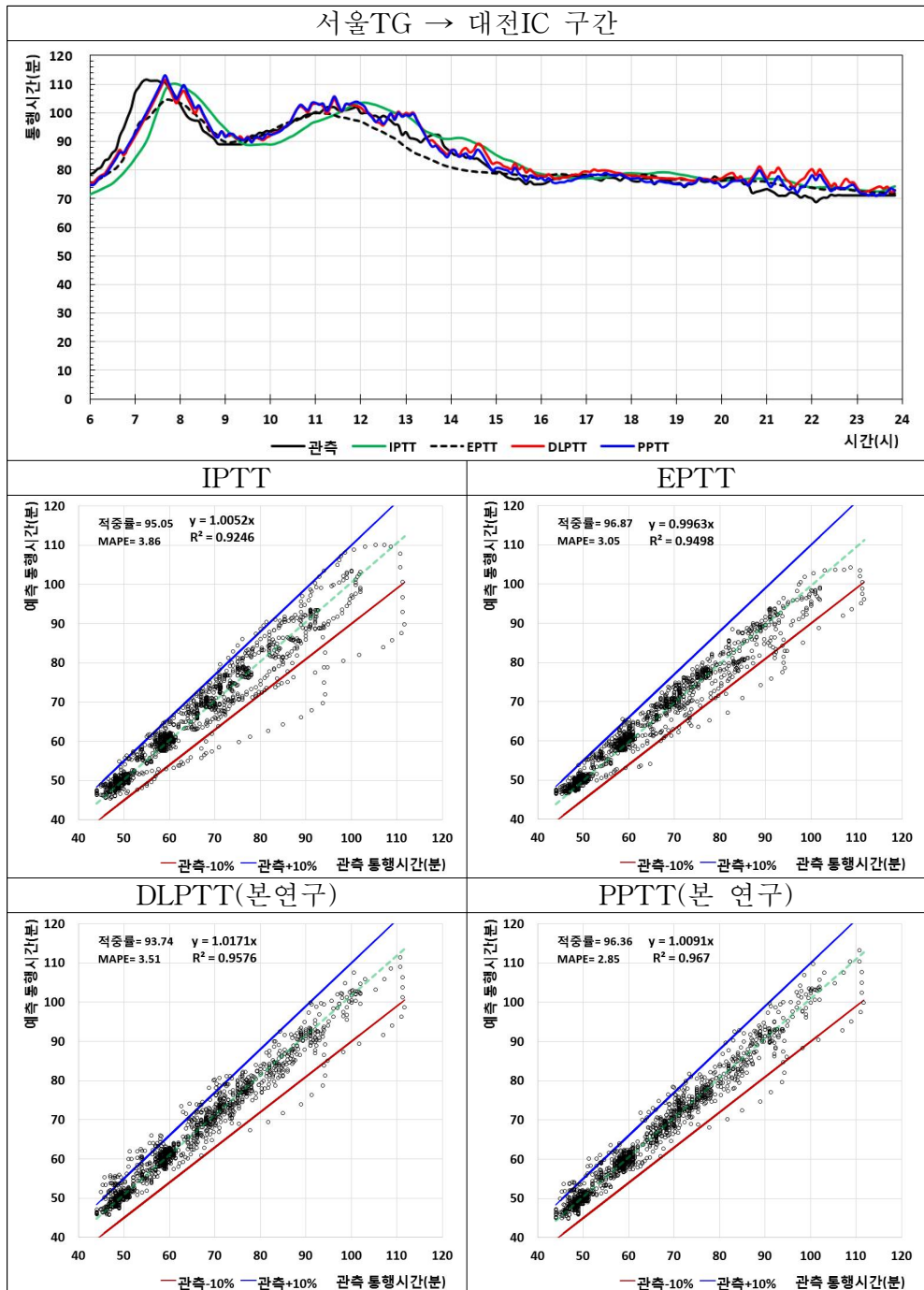
8) 적용 결과: 2봉

분리된 2봉형 통행시간 형태는 오전/오후 첨두를 보이는 경우 전형적인 혼잡-비혼잡-혼잡의 과정을 통해 나타난다. 그러나 <그림 5-51>과 같은 연속적인 2봉형은 혼잡-완화-혼잡의 형태로 나타나며 Stationary jam(SJ)의 꼬리가 고정된 상태에서 머리가 성장-소멸을 반복하며 형성된다. 개발모형은 예측시점 6시를 기준으로 약 1시간 후의 혼잡발생을 예측함과 더불어 혼잡발생 1시간 이후의 혼잡완화 후 혼잡의 재성장을 예측하고 있다. 시점 8시를 기준으로 6시간 동안 혼잡의 지속을 예측하고 있으며, 예측시점 12시를 기준으로 약 4시간 후 혼잡소멸을 예측하고 있다.

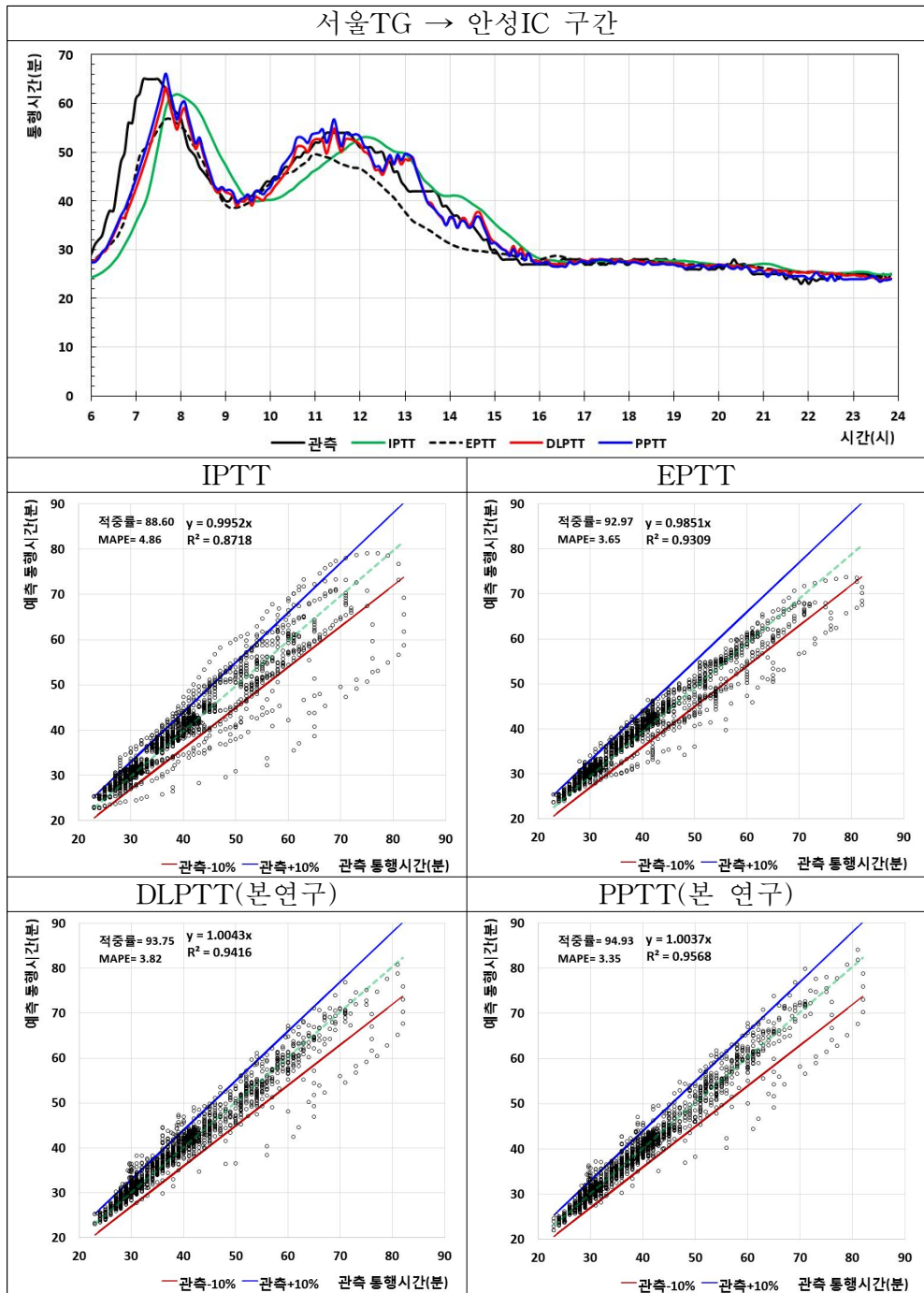
따라서 개발모형의 예측오차(%)는 <그림 5-52>에서 각각 3.51, 2.85로 나타나고 있으며 $\pm 10\%$ 이내 적중률(%)은 93.74, 96.36으로 높게 나타나고 있다. IPTT와 EPTT 모두 95% 이상의 적중률을 보이고 있지만, 이는 통행의 출발점에서 혼잡이 가까우며, 2봉 혼잡의 본체인 SJ가 지소적으로 유지되면서 통행시간의 변동량이 크지 않기 때문이다.



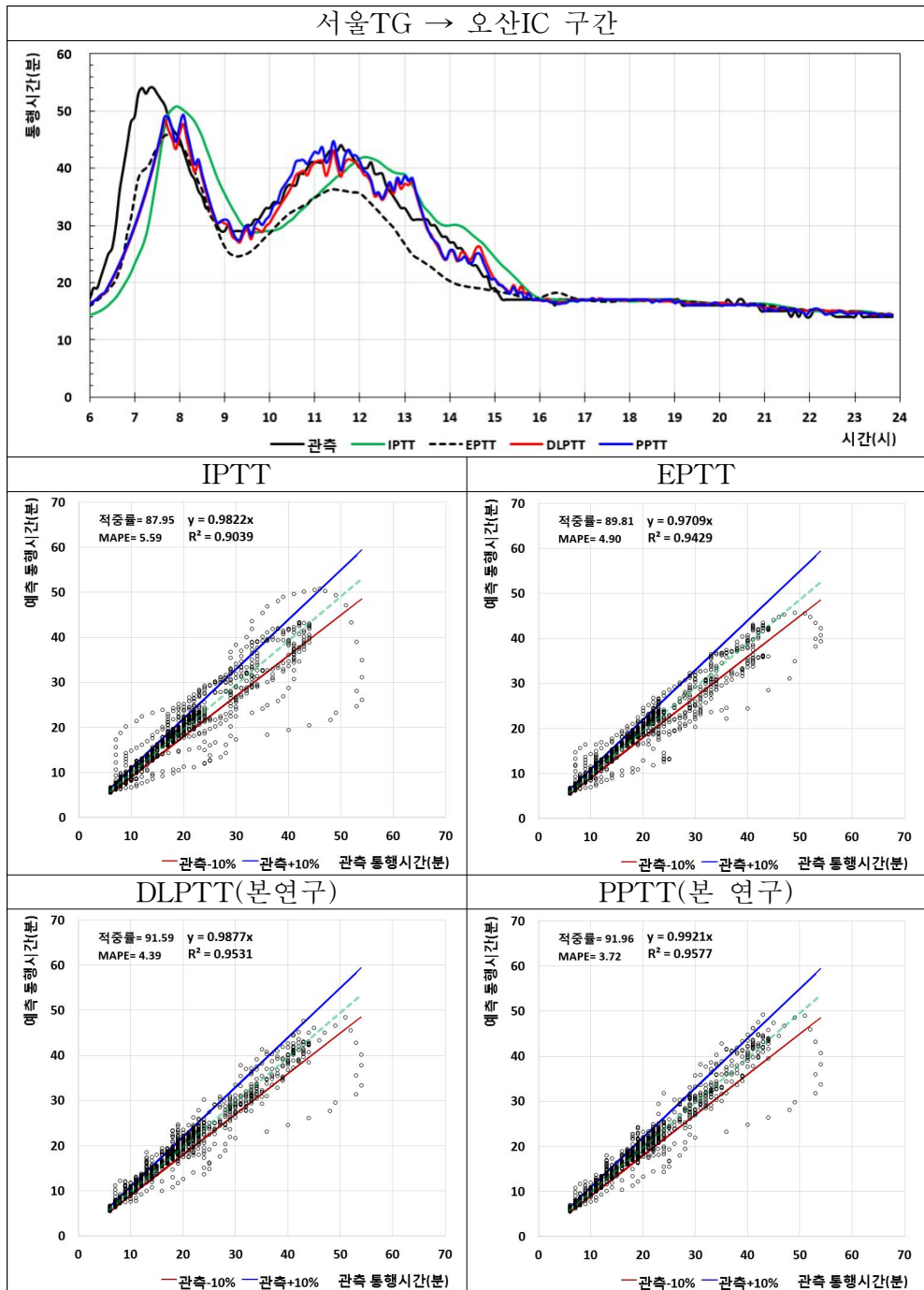
<그림 5-51> 소통상태 예측결과: 2봉



<그림 5-52> 장거리 경로통행시간 예측결과: 2봉

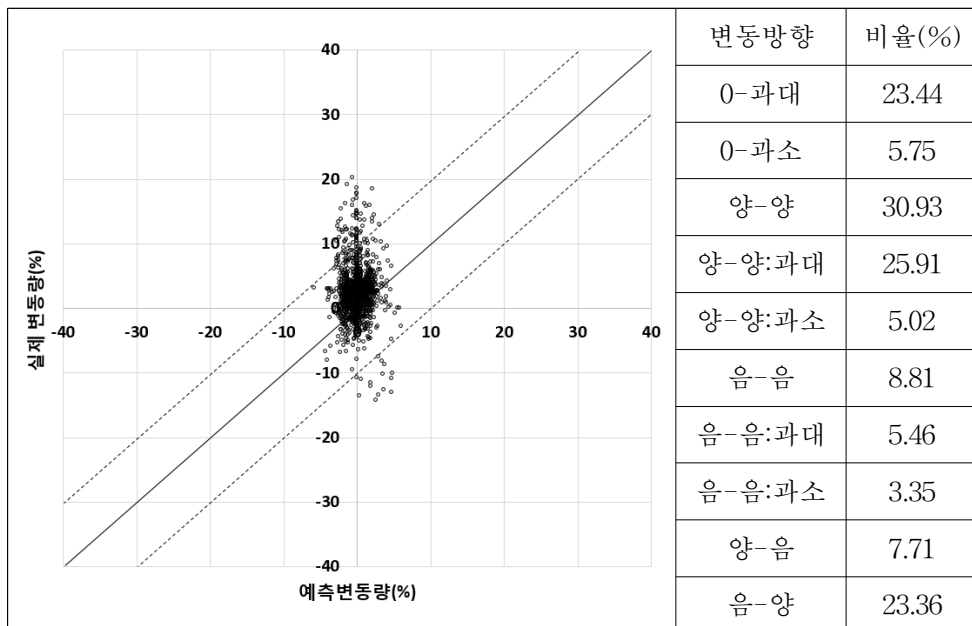


<그림 5-53> 중거리 경로통행시간 예측결과: 2봉

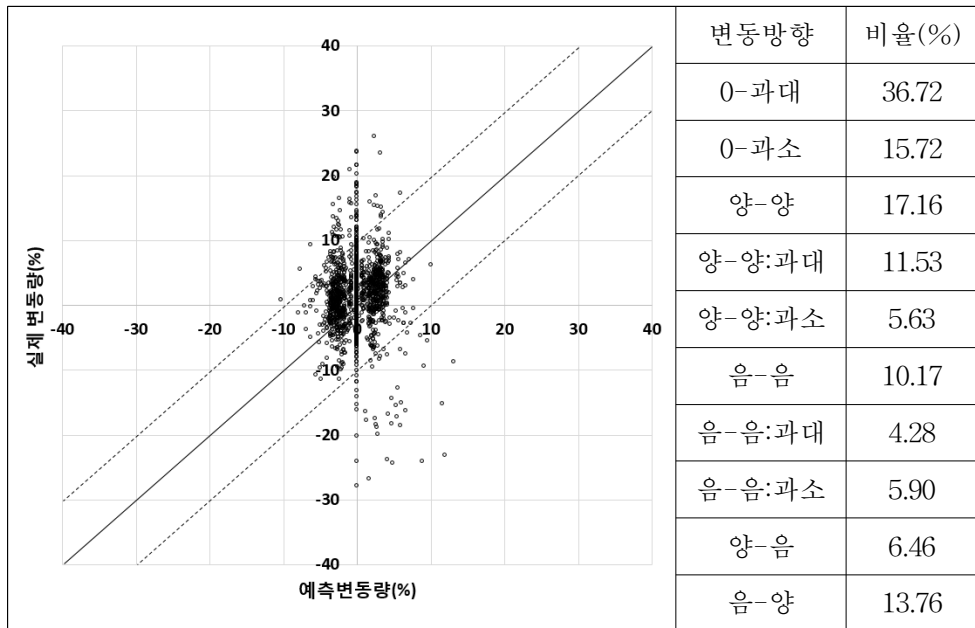


<그림 5-54> 단거리 경로통행시간 예측결과: 2봉

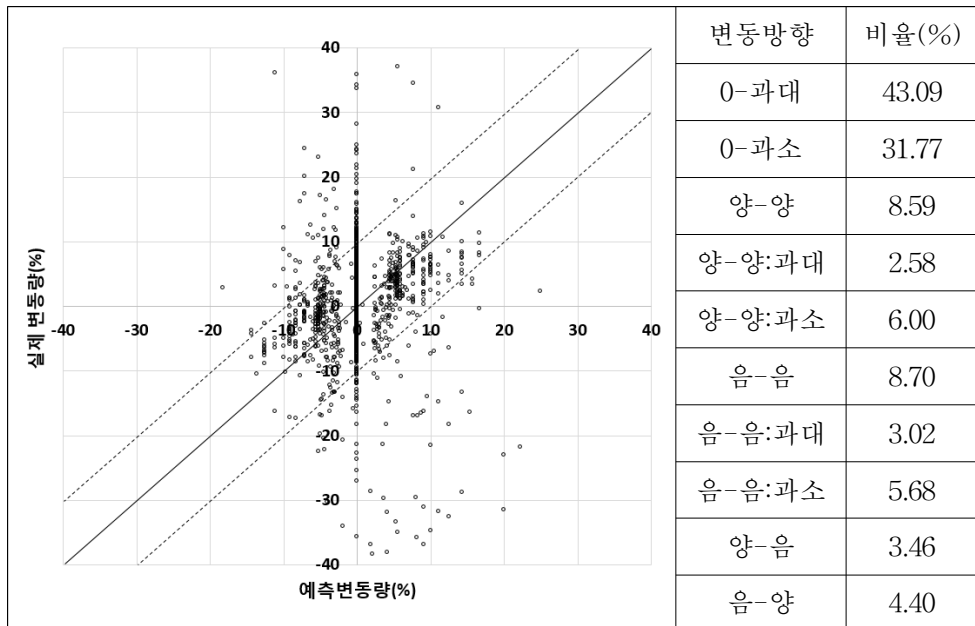
단·중·장거리 통행시간의 $\pm 10\%$ 이내 예측 변동량은 91.67%, 93.84%, 93.89%로 각각 나타났다. 장거리 통행시간의 예측 변동량 구성비(%)는 0-과대 23.44, 양-양 30.93, 음-양 23.36로서 총 77.73%가 양의 방향으로 편중되었다. 그러나 0-과대 29.19%, 완전 예측선 방향 31.37%, 역방향 31.07%로 나타나 특정 방향성에 대한 편중은 보이지 않고 있다. 그리고 변동량은 +3%에 집중되어 있기 때문에 예측오차는 3.51% 이하 적중률은 93.74%로서 예측력은 우수하다고 판단된다. 중거리 통행시간의 경우 실제 변동량이 -3%, 0%, +3%에서 3개 그룹으로 구분된 것 이외에는 장거리 통행시간과 유사한 패턴을 보이고 있다. 따라서 예측오차는 3.82% 이하 적중률은 93.75% 이상으로 나타났다. 단거리 통행시간의 경우 변동의 형태는 타원형으로 보이지만, 0변동의 구성비 74.86%을 제외하면 완전 예측선의 $\pm 10\%$ 이내에 대부분이 분포하고 있다. 양-양 8.59% 음-음 8.70%로서 완전 예측선 방향의 변동은 17.29%로 나타나 역방향 7.86%보다 높게 분석되고 있다. 따라서 예측오차는 4.39% 이하 적중률은 91.59% 이상으로 나타나 예측력은 우수한 편이라고 판단된다.



<그림 5-55> 장거리 경로통행시간 예측 변동분석: 2봉



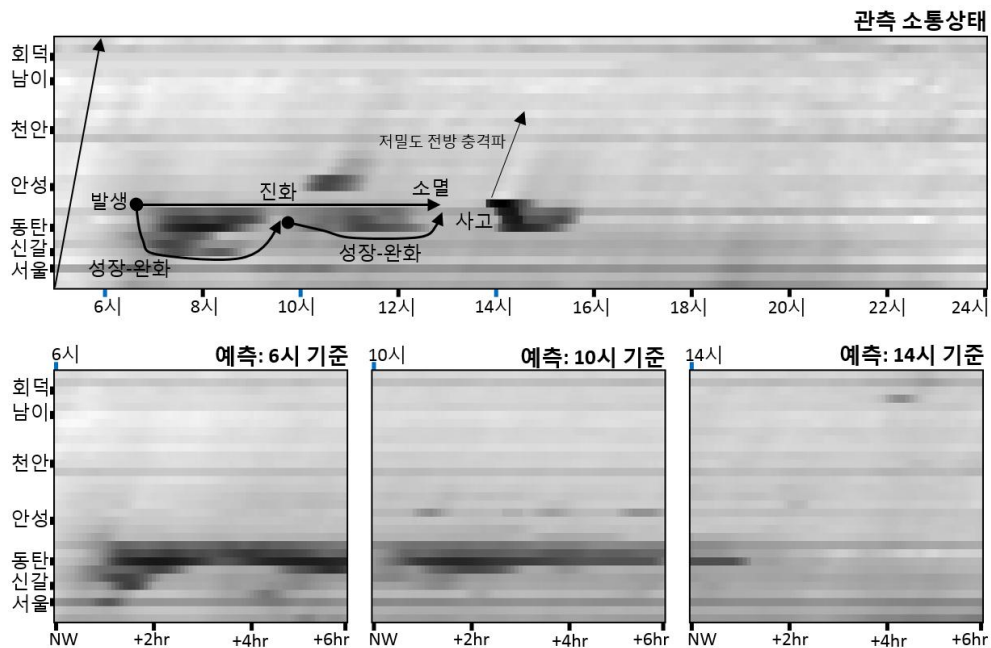
<그림 5-56> 중거리 경로통행시간 예측 변동분석: 2봉



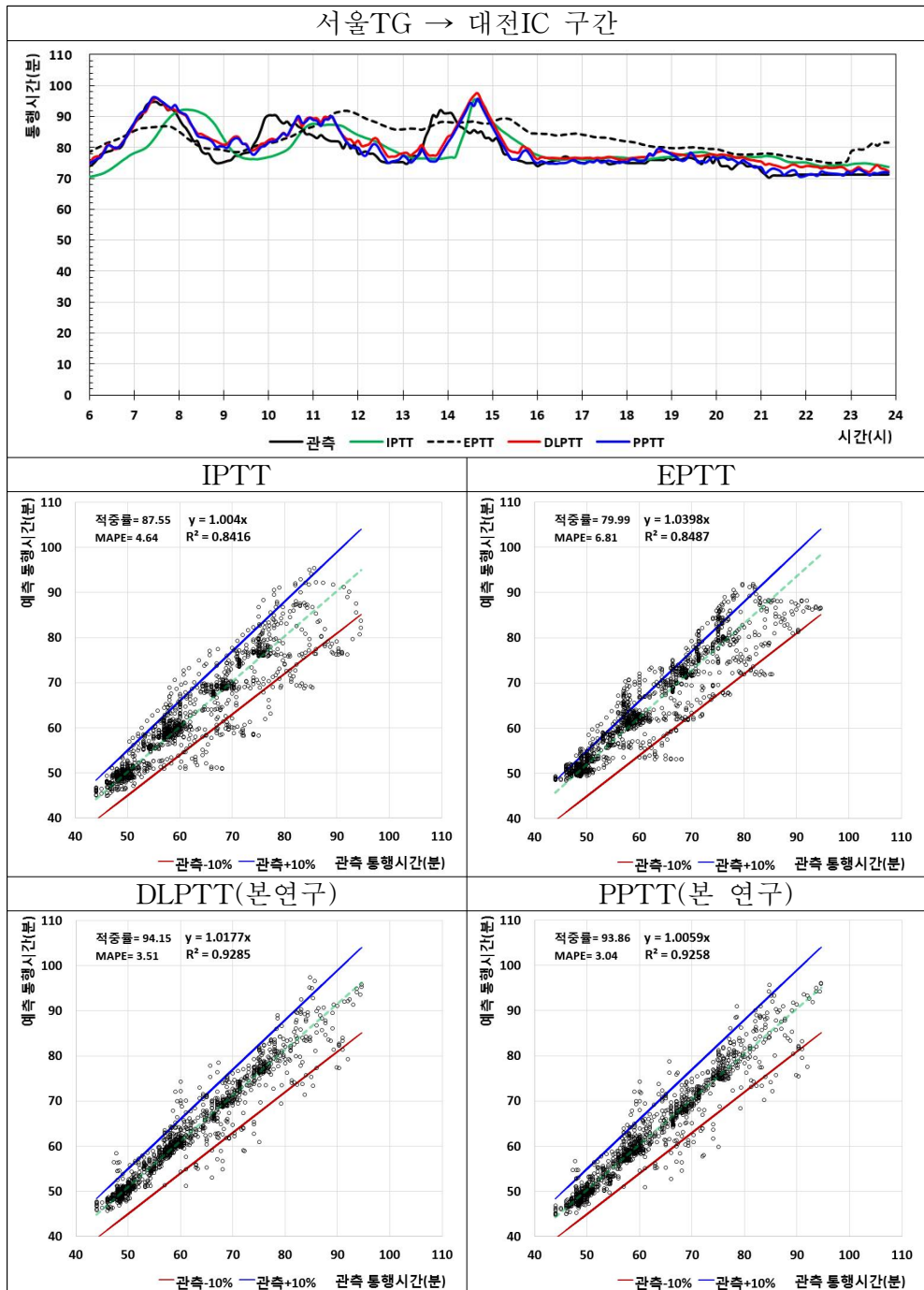
<그림 5-57> 단거리 경로통행시간 예측 변동분석: 2봉

9) 적용 결과: 3봉

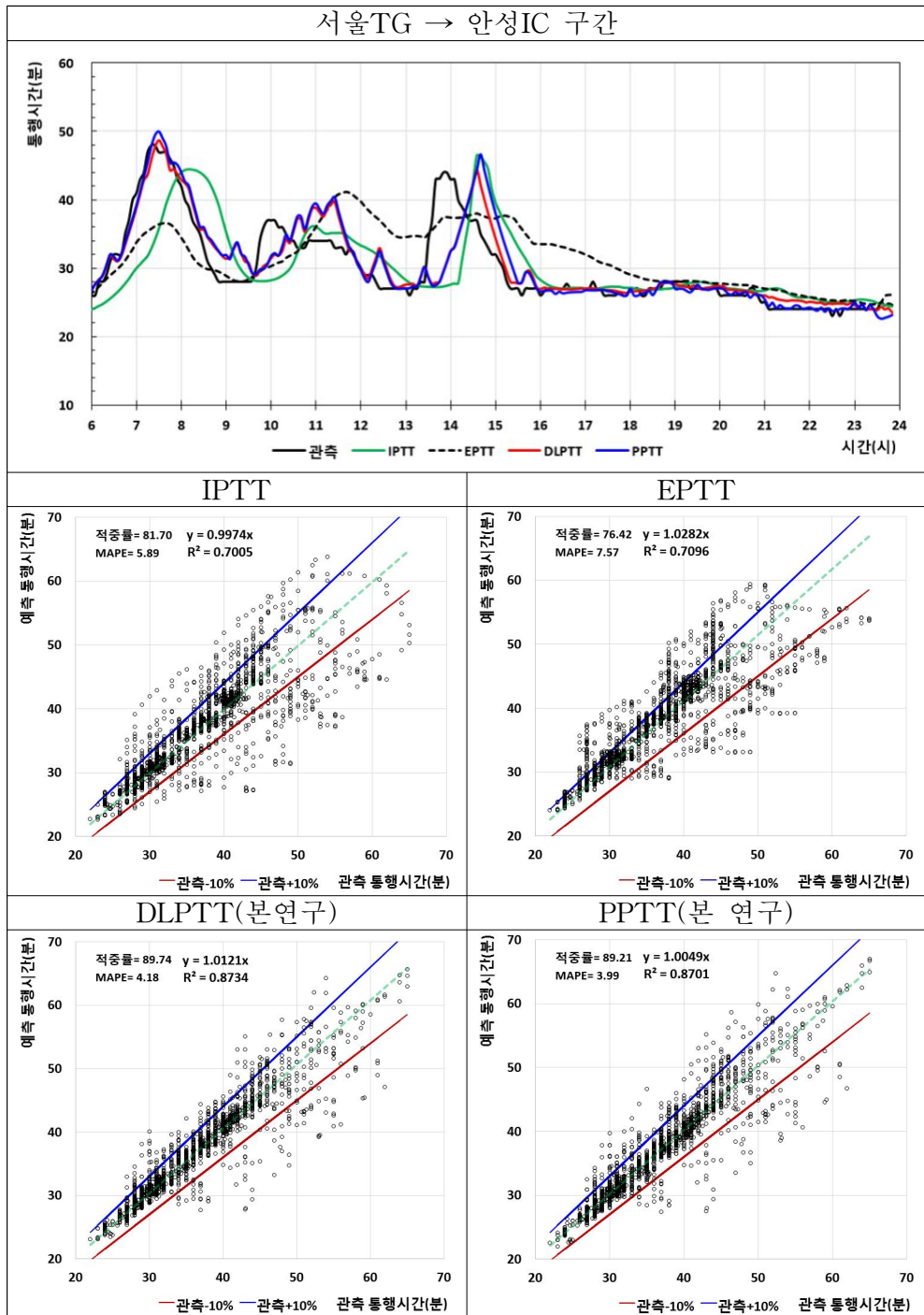
3봉 형태의 혼잡은 관측하기 어려운 형태중 하나이며, 2봉 형태와 동일한 과정을 통해 형성되며, <그림 5-58>과 같다. 그리고 통행시간 예측모형의 평가를 위한 높은 난이도의 예라고 할 수 있다. 예측시점 6시 기준으로 1시간 후 1번째 혼잡 발생과 혼잡발생 2시간 후 혼잡 완화를 예측하고 있다. 그리고 10시를 기준으로 다시 2시간 이후 혼잡 발생을 예측하고 있으나 14시의 혼잡발생은 설명하지 못하고 있다. 이는 14시에 발생한 혼잡이 교통사고로 인한 혼잡이기 때문이며, 교통사고로 인한 낮은 밀도의 높은 속도로 전파되는 전방충격파를 흰색공간으로 확인할 수 있다. <그림 5-59>에서 개발모형은 20분 정도의 상태회복 시간이 요구되었지만 IPTT보다 빠르게 상태의 방향성을 회복하고 있다. 따라서 개발모형의 예측오차는 3.15% 이하, 적중률은 93.86%로서 비교모형에 비하여 우수한 예측력을 보이고 있다. 추가로 충분한 이력자료가 확보되면 교통사고와 같은 비 반복적 행태의 발생시 장래 상태의 예측에 적용이 가능함을 간접적으로 시사해준다고 하겠다.



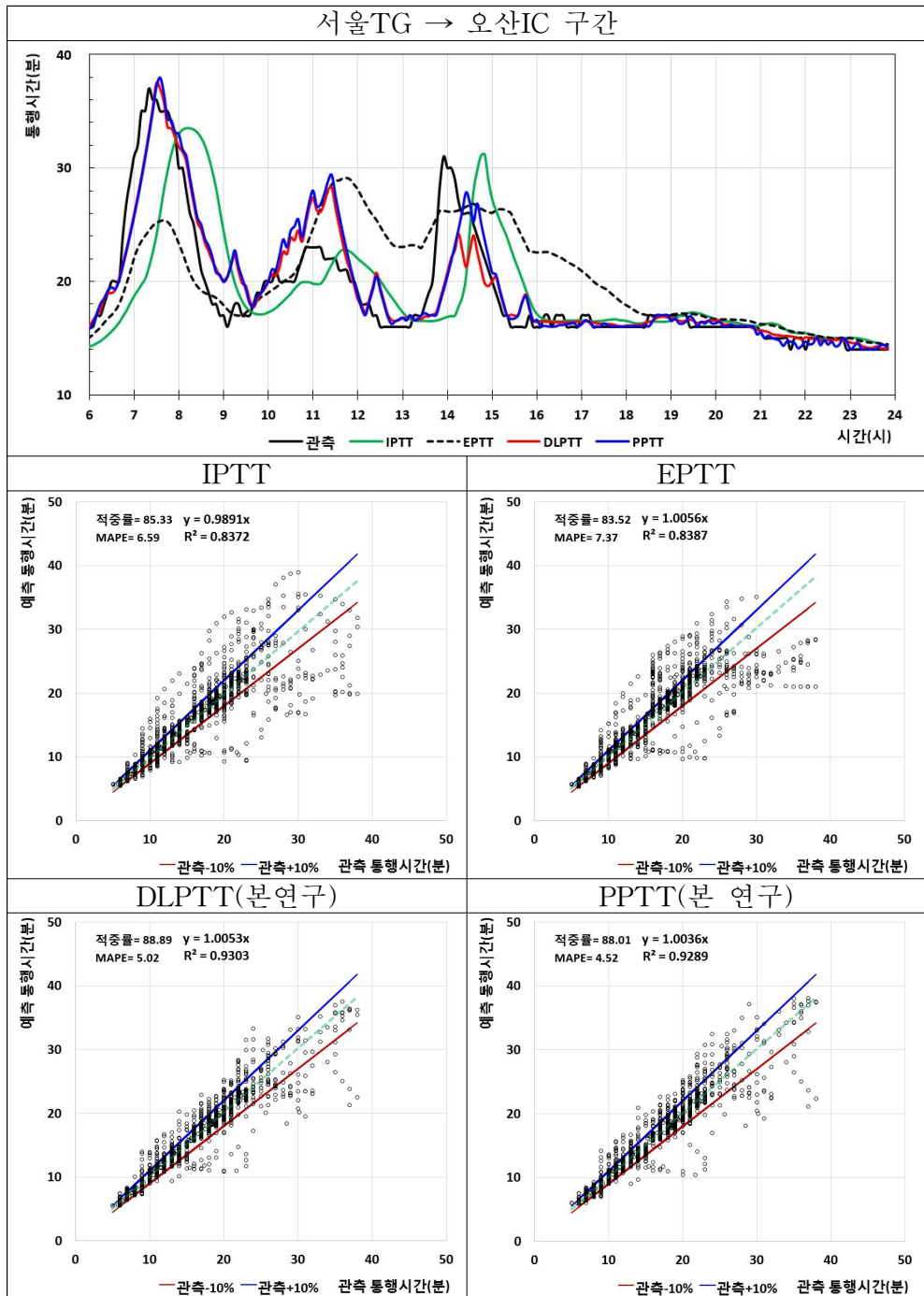
<그림 5-58> 소통상태 예측결과: 3봉



<그림 5-59> 장거리 경로통행시간 예측결과: 3봉

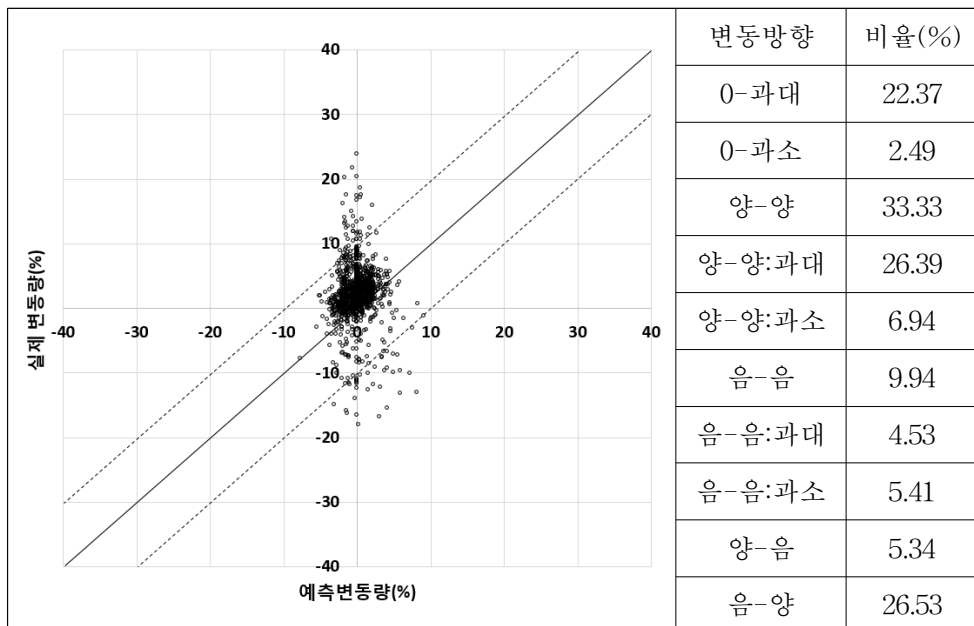


<그림 5-60> 중거리 경로통행시간 예측결과: 3봉

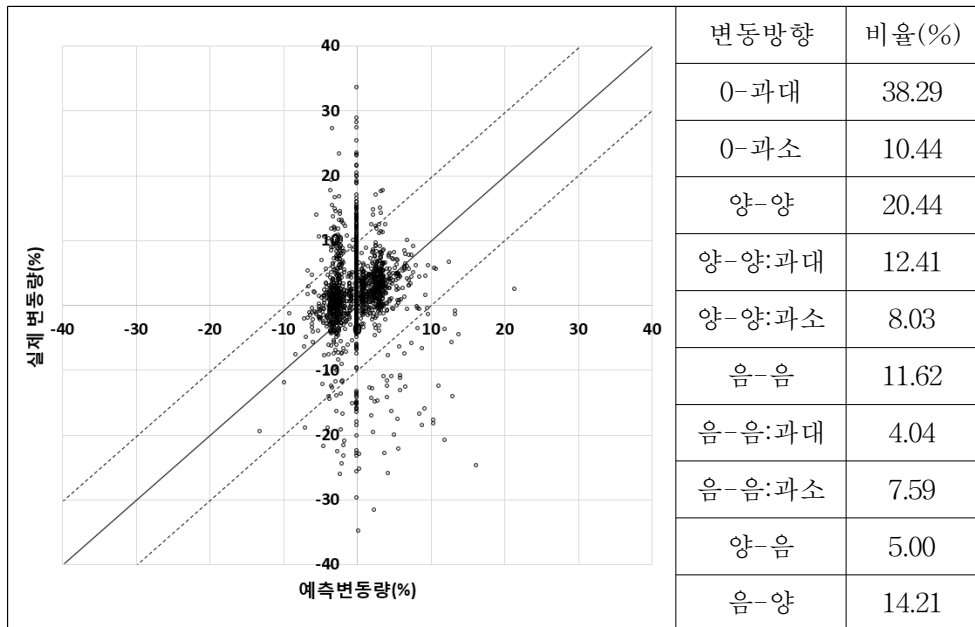


<그림 5-61> 단거리 경로통행시간 예측결과: 3봉

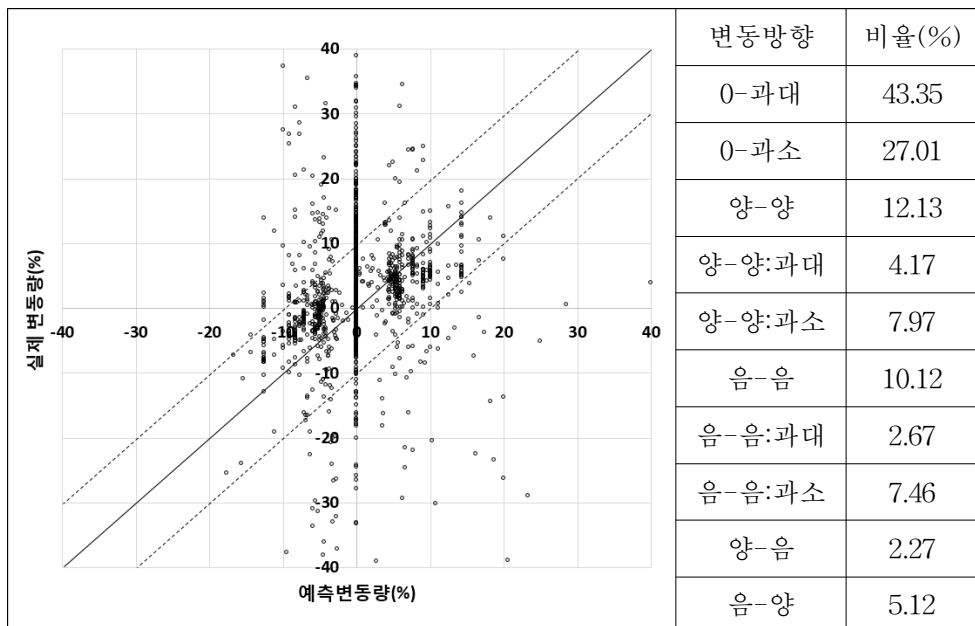
장거리 통행시간의 변동 형태는 집중되어 있으면서 예측 변동량의 +방향으로 편중되어 있다. 0-과대 22.37%, 양-양 33.33%, 음-양 26.53%로 나타나 예측 변동량 +방향의 구성비는 82.23로 매우 높게 나타났다. 그리고 양-양 과대는 26.39%로 양-양 과소 6.94%에 비하여 4배 정도 편중된 과대 예측을 보이고 있다. 그러나 집중된 변동형태의 중심점이 완전 예측선에 가깝게 위치하고 있기 때문에 예측력은 우수하다고 판단된다. 중거리 통행시간의 변동 형태는 실제 변동량 -3%, 0%, +3%에서 3개 형태를 보이고 있으며, -3%와 0%의 형태는 세로로 분포하고 있다. 반면 분포의 중심이 완전 예측선에 위치하면서 예측력은 예측오차 4.18% 이하 적중률은 89.21%로서 우수한 편으로 판단된다. 단거리 통행시간의 실제 변동량 0%의 구성비는 70.36%로서 단거리 통행의 경우 변동이 낮은 것으로 판단된다. 완전 예측선 방향의 양-양 12.13% 음-음 10.12%로서 역변동 2.27%와 5.12%에 비하여 완전 예측선 방향이 높게 나타나고 있으며, 절대 예측 변동량이 큰 경우가 산발적으로 분포하면서 예측오차는 5.02% 적중률은 88.96%로 나타나 예측력은 다소 저하되고 있다고 판단된다.



<그림 5-62> 장거리 경로통행시간 예측 변동분석: 3봉



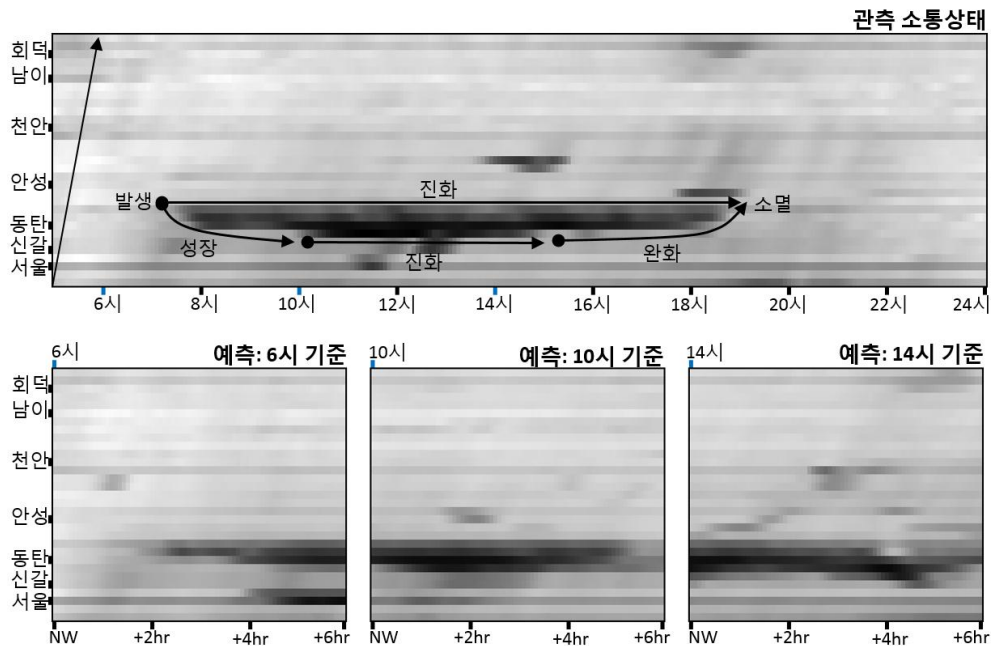
<그림 5-63> 중거리 경로통행시간 예측 변동분석: 3봉



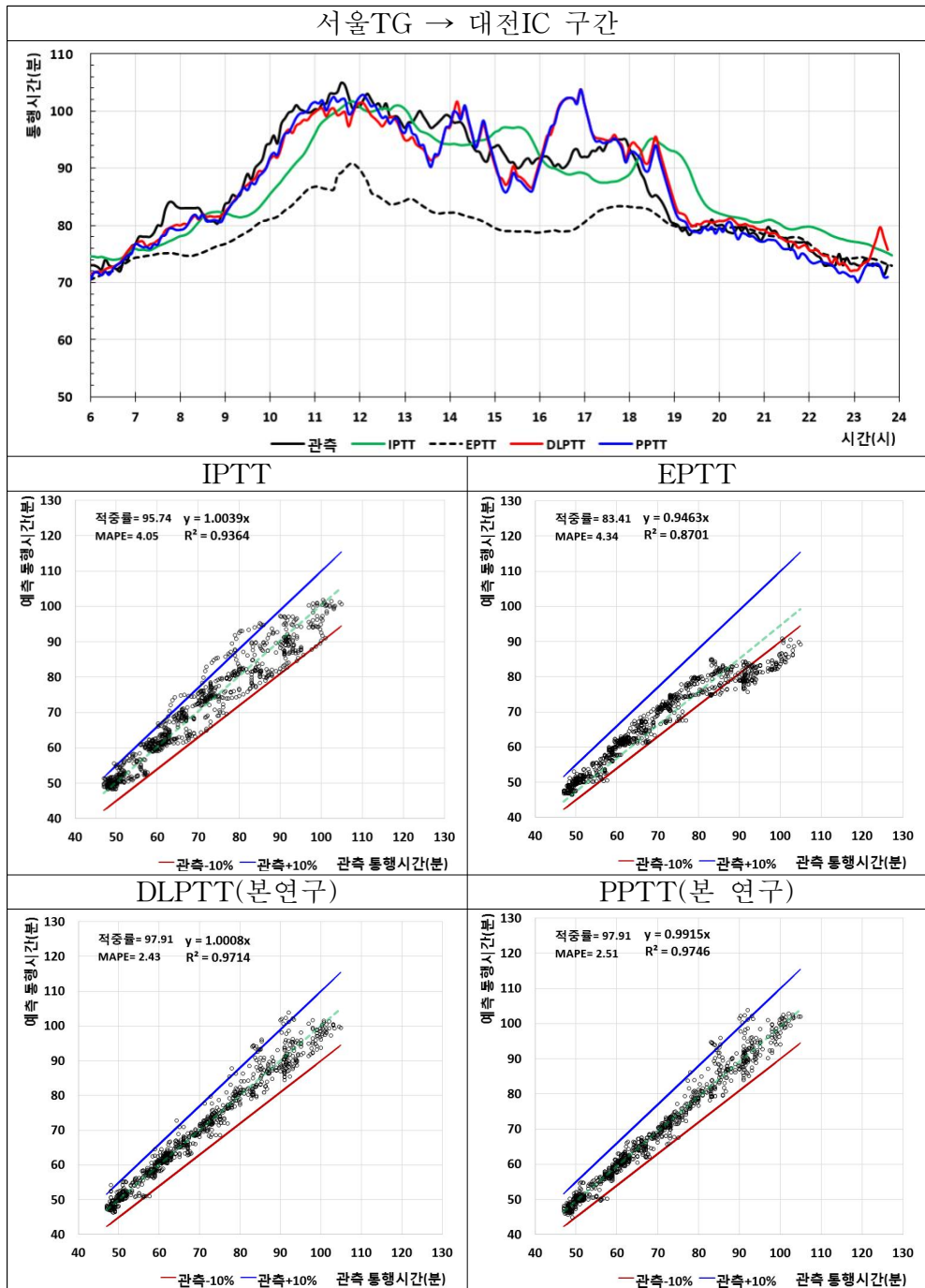
<그림 5-64> 단거리 경로통행시간 예측 변동분석: 3봉

10) 적용 결과: 주간 지속형

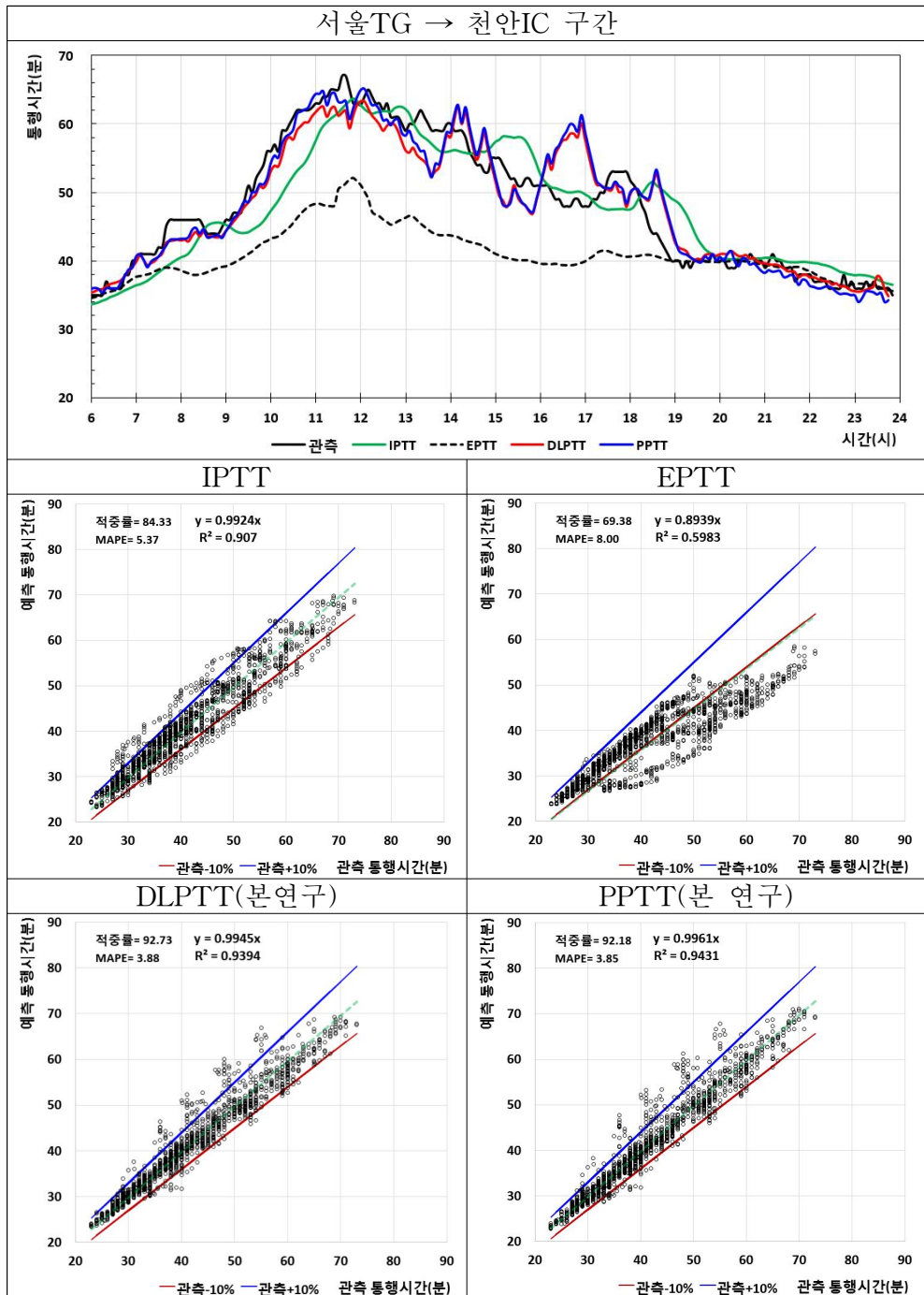
주간 지속형 혼잡은 Stationary jam이 형성된후 장시간 동안 시간적 진화를 진행하면서 혼잡의 머리가 공간적 진화를 앞/뒤로 반복한 후 서서히 완화되는 과정을 거치게 되며, <그림 5-65>와 같다. 혼잡은 7시30분에 시작하여 19시까지 11시간 30분간 장시간 지속되었다. 개발모형은 예측시점 6시 기준으로 2시간 이후부터 혼잡발생을 예측한 후, 예측시점 10시를 기준으로 다시 6시간 동안의 지속적 혼잡을 예측하고 있다. 그리고 시점 14시를 기준으로 5시간 이후의 혼잡완화를 정확히 예측하고 있다. 따라서 통행시간 예측결과의 적중률은 <그림 5-66>에서와 같이 약 98%로서 IPTT와 EPTT의 95.74%와 83.41%보다 높게 나타나고 있다. IPTT의 경우 혼잡의 시간적 진화과정에서 발생한 혼잡의 강도차이로 발생하는 통행시간의 방향성을 설명함에 있어 지속적인 실패를 보이고 있다. EPTT 또한 90분 이후부터 예측능력을 상실하고 있는 것으로 판단되며, 이는 전일 혼잡이 반복적인 단일 패턴이 아니며 패턴과 패턴이 연계된 혼잡이라는 간접적인 의미이기도 하다.



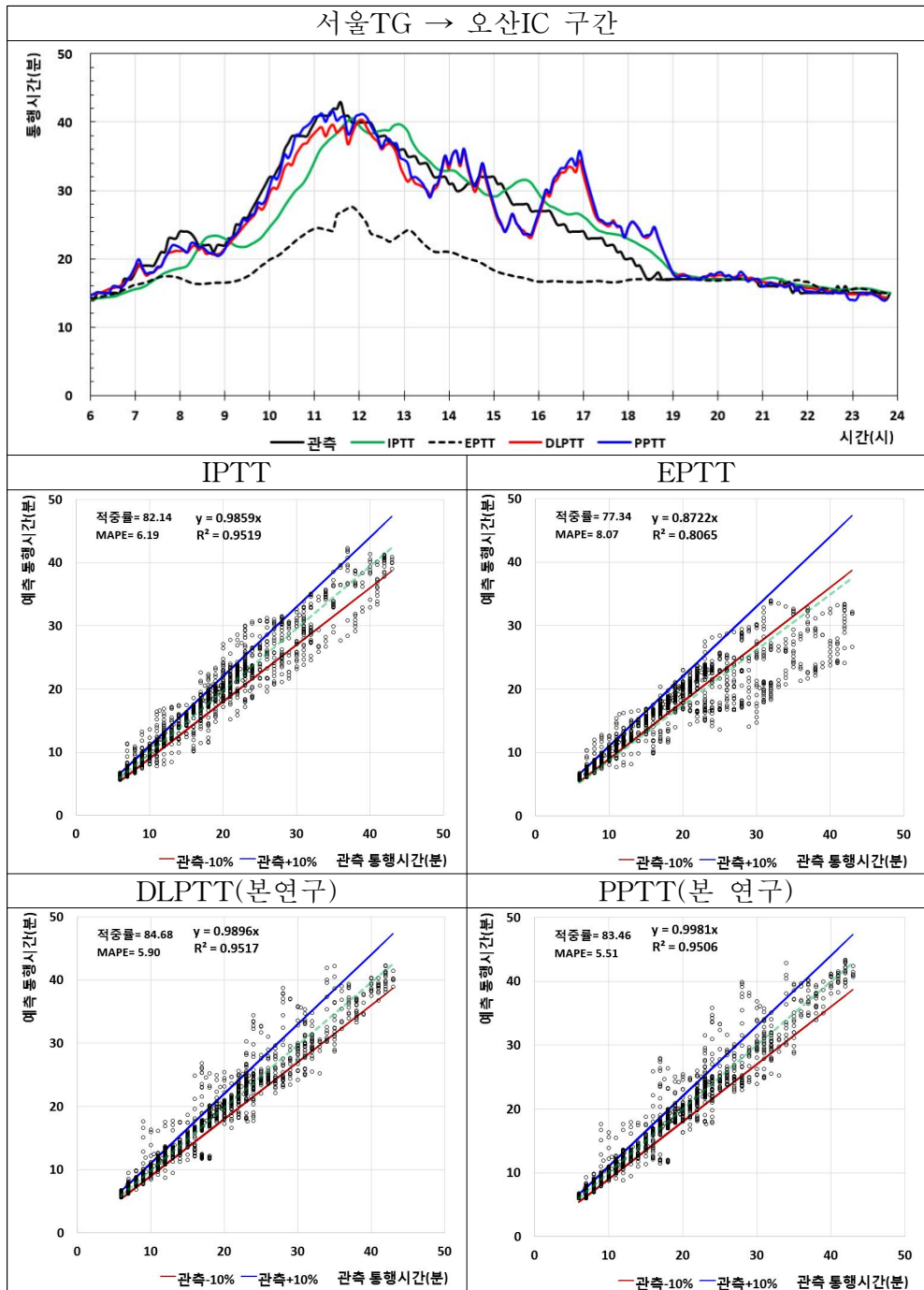
<그림 5-65> 소통상태 예측결과: 주간 지속형



<그림 5-66> 장거리 경로통행시간 예측결과: 주간 지속형

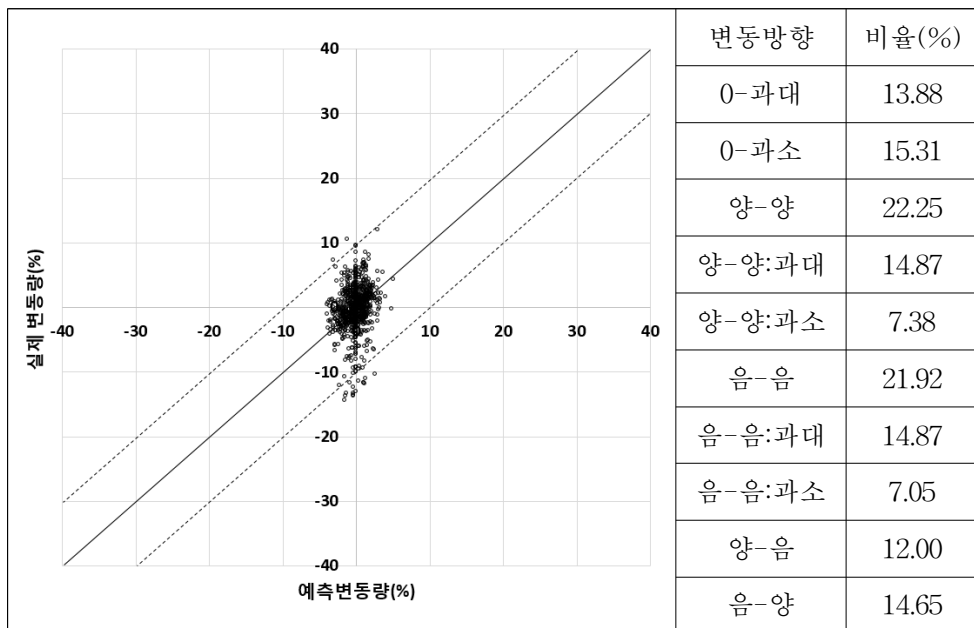


<그림 5-67> 중거리 경로통행시간 예측결과: 주간 지속형

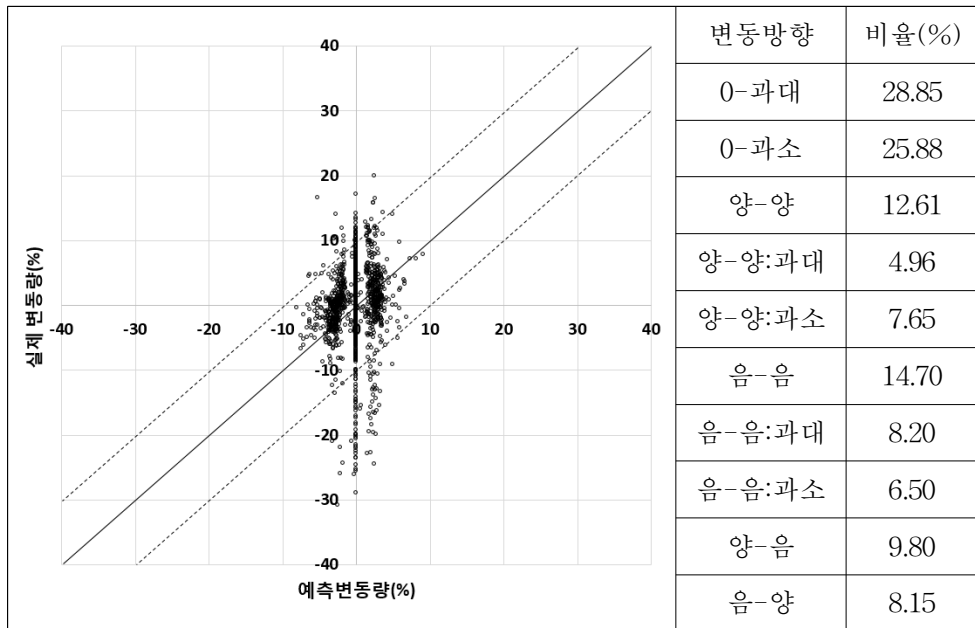


<그림 5-68> 단거리 경로통행시간 예측결과: 주간 지속형

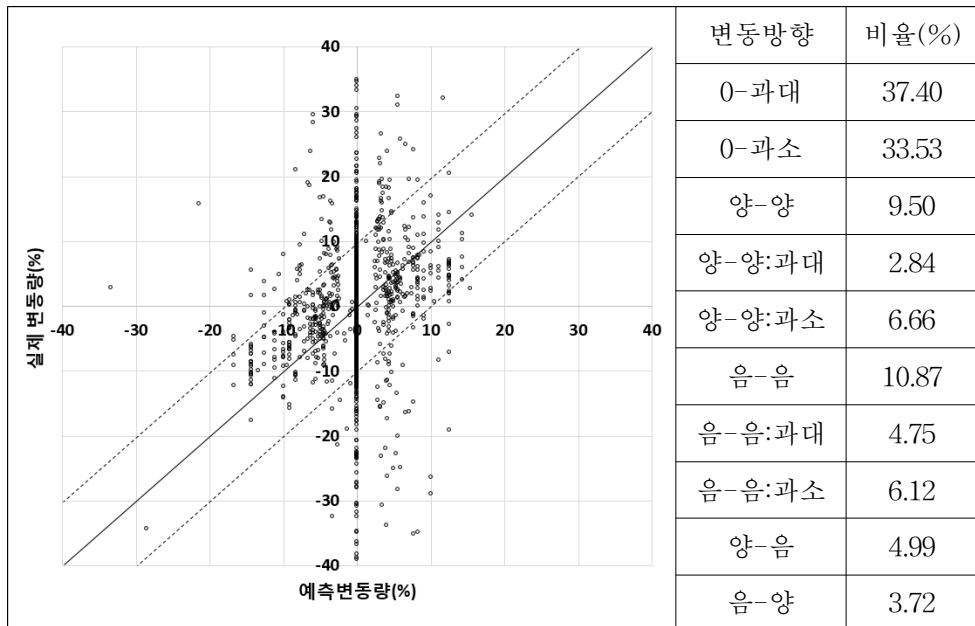
장거리 통행시간의 변동 형태는 원점주위의 완전 예측선 $\pm 5\%$ 이내에 조밀하게 집중되어 있다. 따라서 완전 예측선의 방향성과 상관없이 예측력은 우수하다고 판단된다. 실제 변동량 0의 구성비 29.19%, 양-양 22.25%, 음-음 21.92%, 역변동 26.65%로서 고르게 분포되어 있다. 중거리 통행시간의 실제 변동량이 0인 구성비는 54.73%로 나타났으며, 0-과대 28.85%, 0-과소 25.88%로서 원점을 중심으로 균등하게 분포하고 있다. 양-양 12.61%, 음-음 14.70%로서 완전 예측선 방향의 구성비는 37.31%로 나타난 반면, 역변동은 양-음 9.80%, 음-양 8.15%로 나타났다. 중거리 통행시간의 특징인 3개 변동군이 형성되어 있으며, 예측오차는 3.88% 이하, 적중률은 92.73% 이상으로 나타나 예측력은 우수하다고 판단된다. 단거리 통행시간의 변동 또한 전형적인 3개 집단을 형성하고 있다. 완전 예측선 방향의 양-양 9.50% 음-음 10.47%과 역방향 양-음 4.99%와 음-양 3.72%이 산발적으로 나타나면서 예측오차는 5.90% 적중률은 86.68%로 저하되고 있으나, IPTT와 EPTT보다는 우수하게 나타났다.



<그림 5-69> 장거리 경로통행시간 예측 변동분석: 주간 지속형



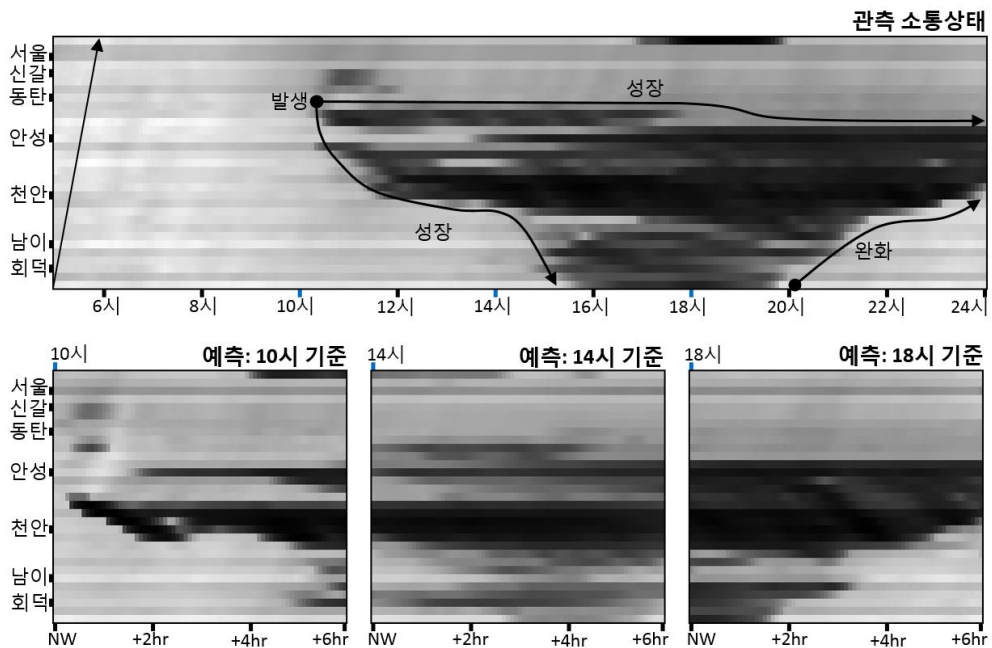
<그림 5-70> 중거리 경로통행시간 예측 변동분석: 주간 지속형



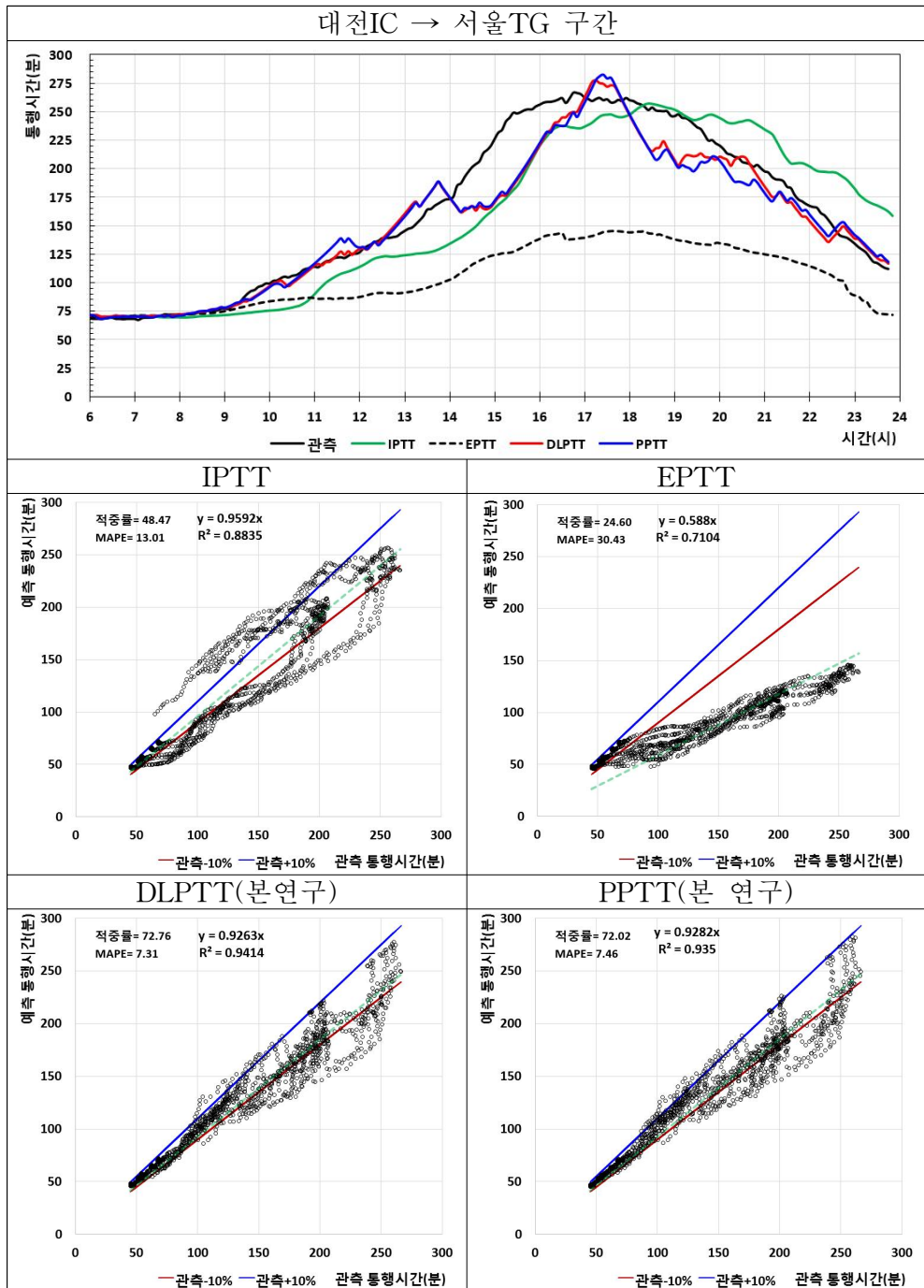
<그림 5-71> 단거리 경로통행시간 예측 변동분석: 주간 지속형

11) 적용 결과: 극심한 정체

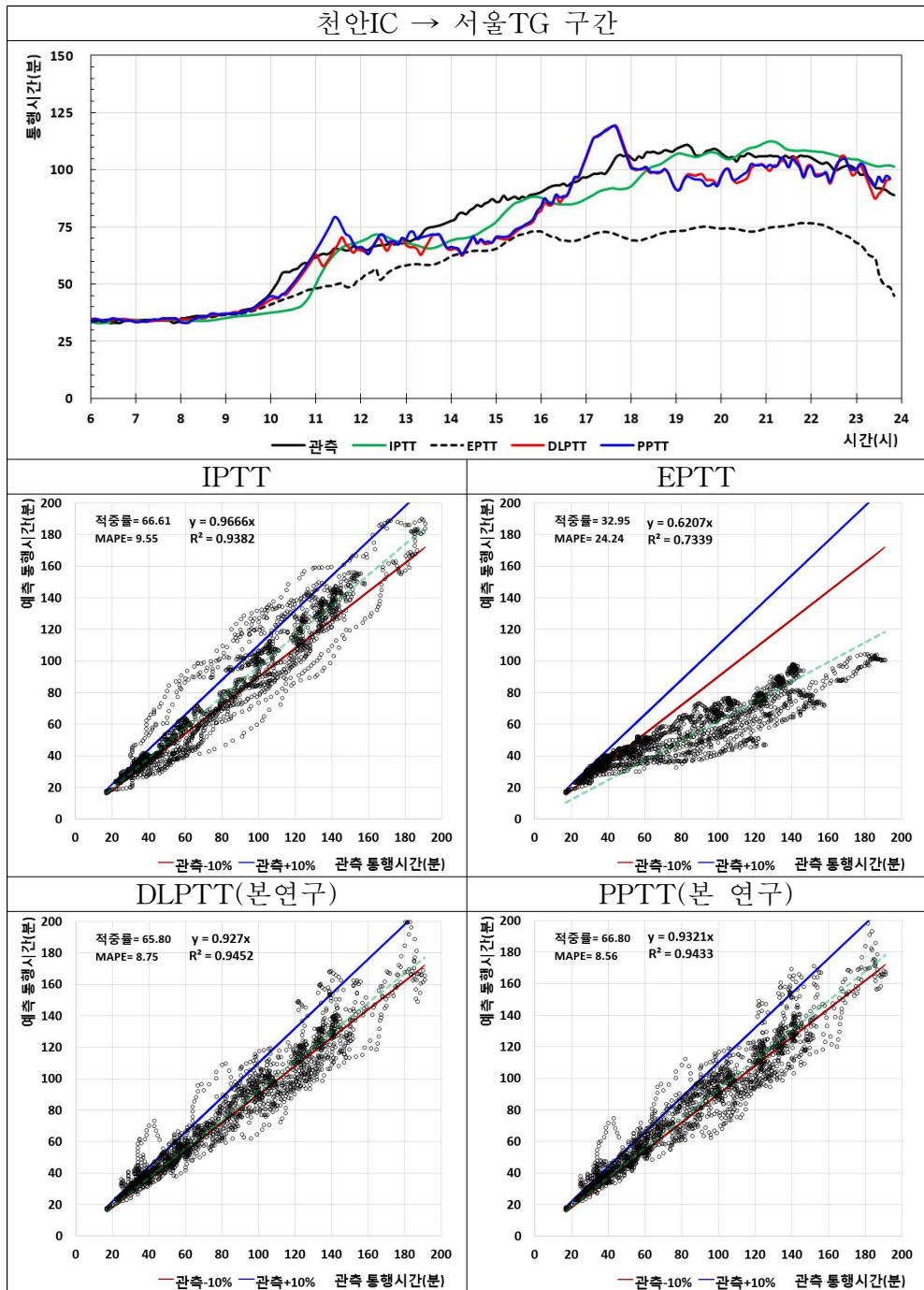
극심한 정체형 통행시간은 거대 혼잡에서 나타난다. 거대 혼잡은 <그림 5-72>와 같이 혼잡발생-빠른 시·공간적 성장-완화-소멸의 과정을 거친다. 혼잡은 동탄 JC에서 대전이남까지 공간적으로 성장하고, 시간적으로 24시 이후까지 지속되었다. 개발모형은 예측시점 10시를 기준으로 혼잡의 시작을 예측함과 더불어 장래 6시간 이후까지 혼잡의 성장을 예측하고 있다. 시점 14시를 기준으로 대전IC 이남까지 혼잡의 공간적 성장을 6시간 전에 예측하고 있으며, 시점 18시를 기준으로 2시간 후 혼잡의 완화가 시작되며 24시이후까지 지속됨을 예측하고 있다. <그림 5-73>의 통행시간 예측결과에서, 도로망의 거대 혼잡은 예측하였지만, 혼잡의 강도에 실패하면서 예측오차는 7.3%이상으로 증가하고 적중률은 72%까지 저하되고 있다. 이는 과거 이력자료의 다양성이 장래 상태를 설명할 만큼 충분하지 않기 때문이며, 충분한 다양성을 확보한 이력자료를 활용할 경우 예측력은 향상 될것으로 판단된다. 반면, IPTT와 EPTT는 예측오차 13.01% 이상, 적중률 48.47% 이하로서 예측실패에 가까운 결과를 보였다.



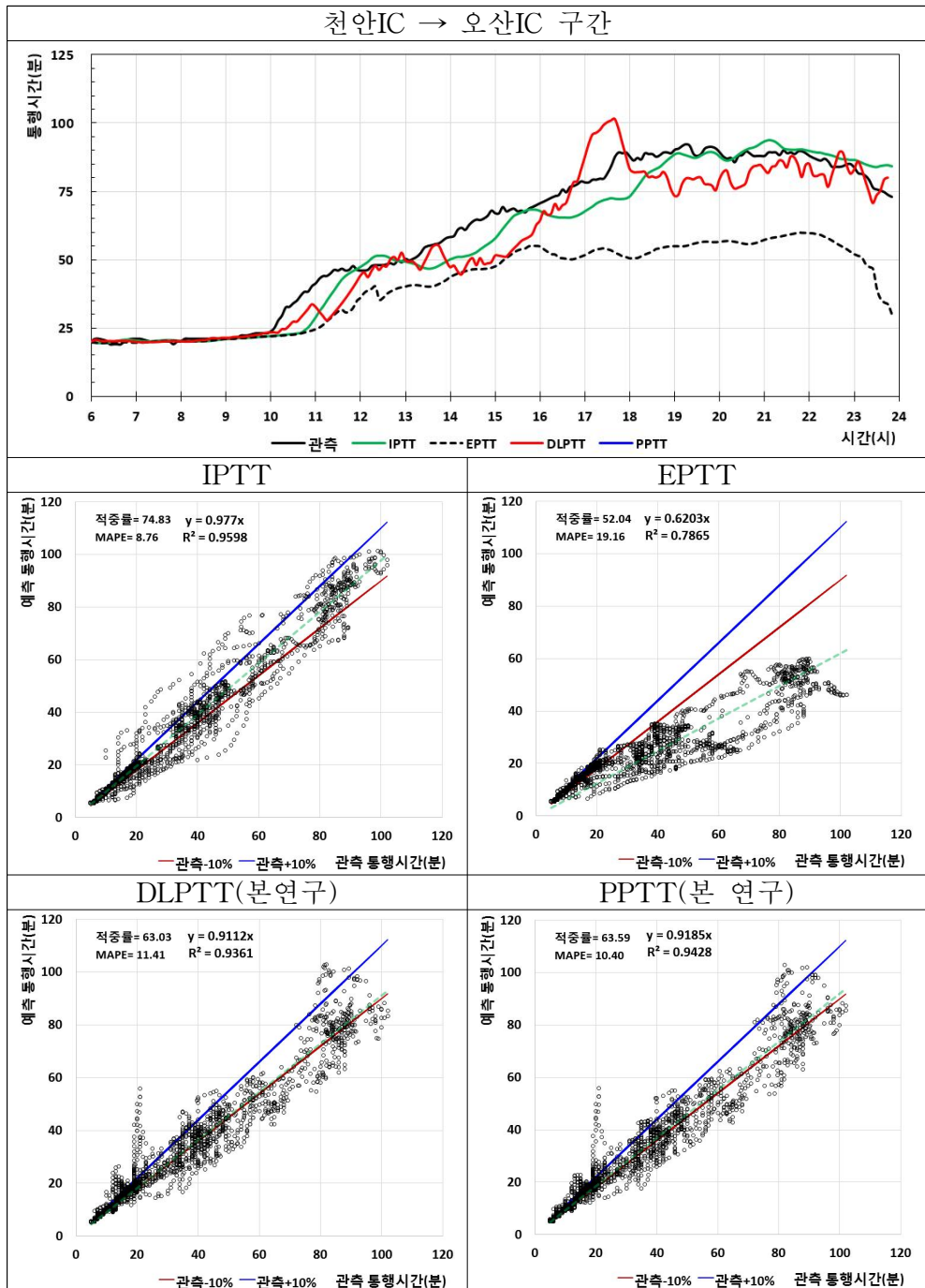
<그림 5-72> 소통상태 예측결과: 극심한 정체



<그림 5-73> 장거리 경로통행시간 예측결과: 극심한 정체

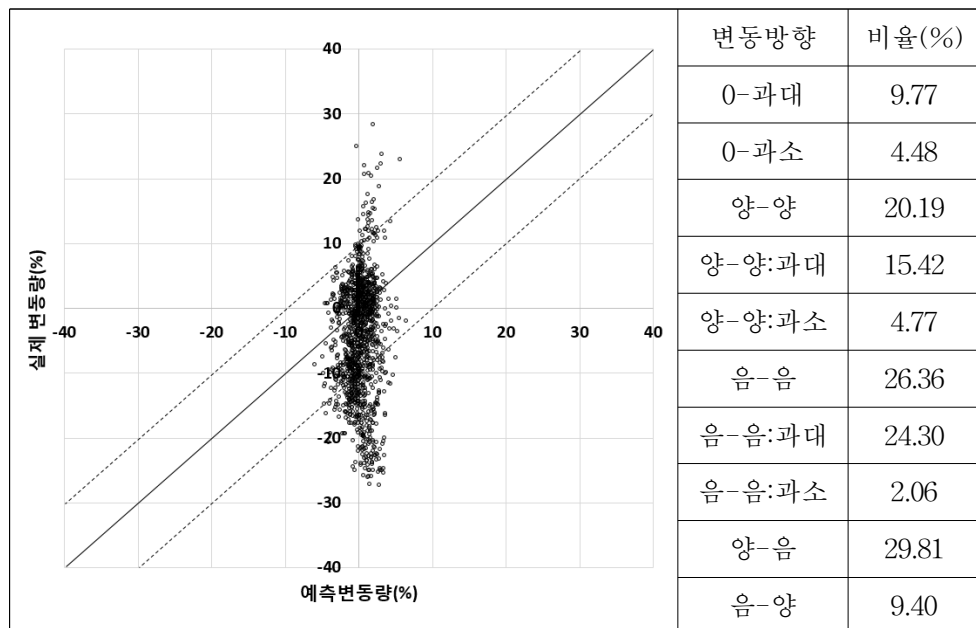


<그림 5-74> 중거리 경로통행시간 예측결과: 극심한 정체

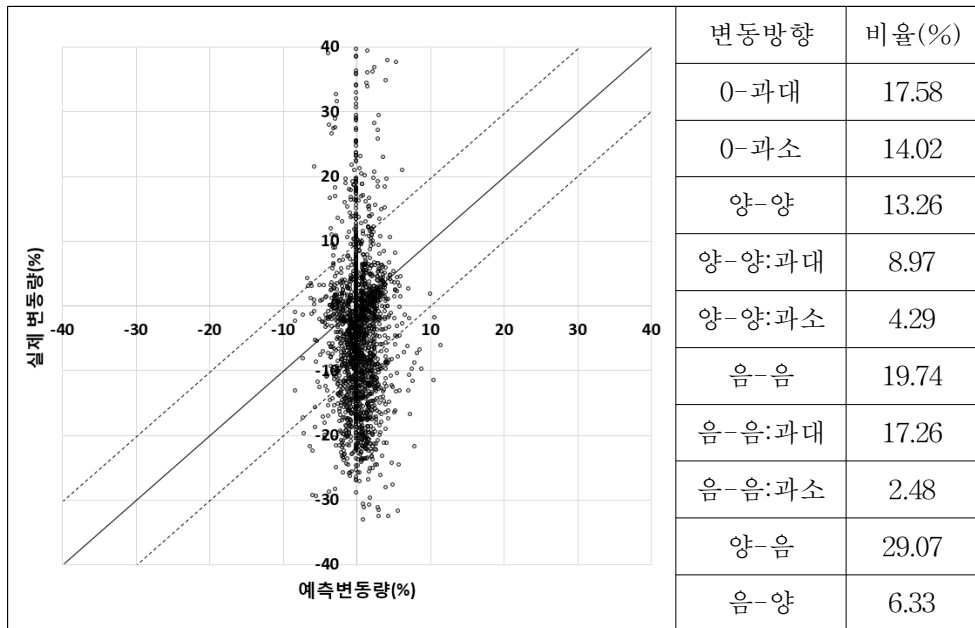


<그림 5-75> 단거리 경로통행시간 예측결과: 극심한 정체

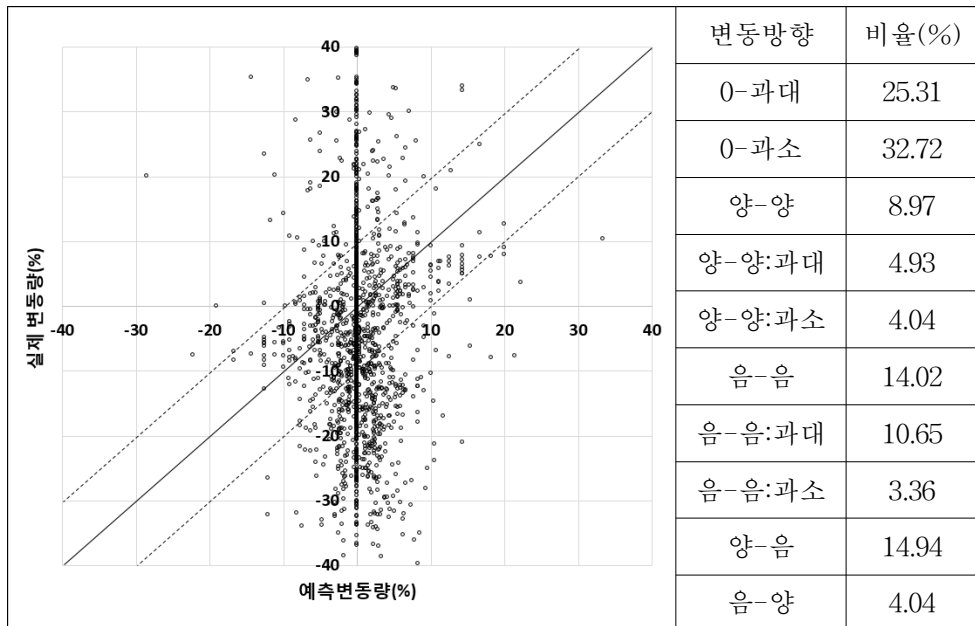
단·중·장거리 통행시간의 변동형태는 예측력이 낮은 경우에 나타나는 세로형을 보이고 있다. 완전 예측선 $\pm 5\%$ 이내의 변동 구성비는 단·중·장거리 통행시간 모두 75%이하로 나타났다. 장거리 통행시간의 양-양 20.19%, 음-음 26.36%로서 완전 예측선 방향은 46.45%로 나타났다. 순변동의 경우 양-양 과대 15.42%, 음-음 과대 24.30%로서 과대 예측의 문제가 발생하고 있다. 역변동의 경우 양-음 29.81%로 나타나 거대 혼잡으로 인한 통행시간 증가량을 과소 예측하고 있는 것으로 나타났다. 그럼에도 불구하고 예측오차는 7.46% 이하 적중률은 72.02% 이상으로 나타났다. 중·단거리 통행시간도 장거리 통행시간의 변동분석 결과와 유사한 결과를 보이고 있다. 특히, 단거리 통행시간의 세로형 변동형태는 넓고 길게 분포하고 있으며, 과소 예측의 경향이 나타나고 있다. 개발모형의 예측오차는 7.46% 이상, 적중률은 65.80% 이하로 감소하고 있지만 예측오차는 10% 이상으로 증가하고 있지 않고 있다. 반면, 비교모형의 경우 <그림 5-73~75>에서 완전 예측실패에 가까운 결과를 보이고 있다.



<그림 5-76> 장거리 경로통행시간 예측 변동분석: 극심한 정체



<그림 5-77> 중거리 경로통행시간 예측 변동분석: 극심한 정체



<그림 5-78> 단거리 경로통행시간 예측 변동분석: 극심한 정체

5. 소결

1) 교통정보 예측에 관한 소결

기존 통행시간 예측의 공간적 영역은 도로구간에 한정되었으며, 도로망의 소통상태 예측은 예측된 도로구간의 집합으로 고려되었다. 단기예측(Short-term forecasting)의 관점에서 볼 때 단일 도로구간의 시계열 정보를 이용한 가까운 장래 상태의 평가는 효율적이라 판단된다. 그러나 기존의 지점검지체계가 구간검지체계로 확장되어 감에 따라 통행시간 정보예측의 측면에서 단기 예측기법은 예측영역의 확장이라는 근본적인 문제에 다시 봉착하였다. 이는 ITS분야의 예측 모델링에 있어 가장 근본적인 문제라 할 수 있으며, 장래 교통상태의 불확실성(Uncertainty)에 기인한다. 결국, 예측은 장래 상태의 불확실성을 최소화하는 문제이기 때문이다. 이상의 한계를 극복하기 위하여, 본 연구에서는 KJC 예측기를 개발하고 다각적인 종합평가를 수행함으로써 개발모형의 성능을 검증하였다. 다각적인 종합평가 결과를 교통정보 예측과 관련된 도로망 소통상태 예측, 경로통행시간 예측, 그리고 패턴인식으로 소결하면 다음과 같다.

본 연구에서는 교통 소통상태의 개념을 기존의 단일 도로구간이라는 제한적 공간영역에서 도로구간의 집합인 도로망으로 확장하였다. 기존 단일 도로구간의 시계열 정보를 이용한 장래상태의 예측은 시간적으로 단기예측의 한계를 벗어날 수 없었다. 다시말해, 수용가능한 수준의 예측 오차라는 측면에서 장래 교통상태의 불확실성 극복은 단기예측 수준에 머물러있었다. 반면, 도로망 기반의 공간적 접근법은 도로구간의 연결체로서 도로망의 전반적인 교통류의 흐름과 통행의 방향성을 고려할 수 있기 때문에 장래 교통상태의 불확실성을 감소시킬 수 있었으며, 예측의 시간적 영역을 중·장기로 확장하였다. 그리고 KJC 예측기는 모델러에 의한 인위적인 접근방식 대신 대용량의 교통 이력자료에 내재된 지식을 이용하도록 개발하였다. 도로망 소통상태의 예측결과, 도로망 기반의 공간적 접근방법은 단기예측의 측면에서 볼 때 단일 도로구간에 대한 국부

적인 예측성능을 저하시킬 수 있으나 수용가능한 수준이었으며, 중·장기 예측의 측면에서 장래 6시간까지 도로망의 소통상태를 예측하였다. 이는 단·중·장기 교통상태의 예측에 있어 단일 도로구간이라는 일개 의사결정자가 아닌 도로망을 구성하는 다수의 도로구간 의사결정자에게 장래 상태를 결정하게 함으로써 장래 상태의 불확실성 문제를 효율적으로 극복할 수 있음을 의미한다. 추가로 본 연구에서 개발된 도로망 기반의 공간적 접근법을 이용하는 KJC 예측기의 공간적 영역은 단일 지점 및 구간에 적용이 가능하며, 예측영역은 6시간 이후까지 확장이 가능하다. 그러나 도시간 경로통행시간 예측이라는 2차 목표를 달성하기 위한 예측영역은 6시간으로 충분하였다.

경로통행시간의 예측 기술은 예측의 정확도 측면에서 볼 때 단거리 통행에 한정되고 있는 실정이다. 물론 과거패턴 등을 이용하여 중·장거리 통행시간 정보를 제공하고 있지만, 정보제공을 위한 불가피한 수단과 예측의 학문적 영역은 구분될 필요가 있다. 본 연구에서는 예측된 도로망의 소통상태를 이용한 동적 구간기반 경로통행시간(DLPTT) 기법과 경로기반 경로통행시간 기법(PPTT)을 적용하여 단·중·장거리 통행시간을 예측하였다. 비교모형으로 선정된 동시 경로통행시간(IPTT) 기법과 경험기반 경로통행시간(EPTT) 기법에 비하여 본 연구에서 개발된 경로통행시간 예측기법은 예측오차와 적중률의 관점에서 볼 때 단·중·장거리 모든 경우에서 기존의 통행시간 예측기법보다 우수한 성능을 보였다. 따라서 본 연구에서 개발된 KJC 예측기는 도시간 경로통행시간을 포함하는 단·중·장거리 통행시간의 예측에 적합하다고 판단된다. 추가로 자료관리시스템 등의 본 연구에서 개발된 KJC 예측기의 운영환경이 조성되지 않는 경우, IPTT는 단거리 통행시간의 영역에 EPTT는 반복적 통행행태가 뚜렷한 도시부에 적합하다고 판단된다.

본 연구에서 개발된 KJC 예측기는 KNN-NPR과 j -군집화가 결합된 강력한 패턴인식을 기반으로 한다. 그리고 도로망 소통상태 예측과 이를 이용한 경로통행시간의 예측에 있어 본 연구에서 개발된 KJC 예측기는 장래 상태의 불확실성 문제를 효과적으로 극복하면서 기존 예측모형에

비하여 매우 우수한 결과를 보였다. 이상의 패턴인식 기반의 모형개발과 예측의 우수성은 방대한 양의 이력자료에는 패턴의 경계가 분명하든 분명하지 않든 국부적 패턴이 존재하며 그 패턴을 이용한 예측문제 접근법은 장래상태의 불확실성을 감소시키는데 효과적인이라는 직접적 증거이다. 그리고 이력자료에 내재된 패턴은 도로망의 소통상태 패턴과 더불어 도로망의 소통상태 패턴과 경로통행시간의 패턴이 동시에 존재한다고 판단된다. 추가로 개발모형의 우수성은 도로망의 상태 진화가 반복적 행태와 비반복적 행태가 공존하는 초기결정성이 강한 혼재상태(Mixed state)에 가깝다는 사실을 간접적으로 말해 준다. 이는 개발 모형이 KNN-NPR을 기반으로 개발되었으며, KNN-NPR은 혼돈상태이론(Chaotic state theory)에 기초하고 있기 때문이다.

2) 모형개발에 관한 소결

단·중·장거리 경로통행시간 정보제공을 위한 도로망의 소통상태를 예측하기 위해서는 대규모 도로망에 적용이 가능하도록 모형의 연산수행속도와 파라미터 최적화에 대한 문제를 해결해야 한다. 개발모형의 수행속도는 단일 구간 기반의 순차적 연산대신 도로망 기반의 일괄적인 연산을 수행하기 때문에 기존의 구간기반 모형보다 매우 빠른 연산수행속도를 보였다. 예를 들어 10만개의 도로구간으로 구성된 도로망이 있을 경우, 구간기반 모형은 10만개의 모든 도로구간에 대한 예측을 수행해야 한다. 따라서 5분 시간간격으로 장래 6시간을 예측할 경우 720만번의 연산을 수행해야 한다. 추가로 이는 기존의 단기예측기법이 장래 6시간까지 예측 정확도를 확보한 경우의 예이다. 그러나 본 연구에서 개발된 모형은 1만개로 구성된 10개 도로망을 구성하고, 10번의 연산으로 10만개 도로구간에 대한 6시간까지의 장래 소통상태를 예측함으로써 연산수행속도가 매우 빠르다. 추가로 KJC 예측기의 빠른 연산수행속도는 동적 교통정보의 생애주기를 연장하는 역할을 수행한다.

고도화된 예측모형의 파라미터 최적화는 모형의 적용에 있어 또 다른 극복대상이 되는 경우가 많으며, 지속적인 파라미터 값의 조정은 모형의

적용성을 저하시키는 요인이 되고 있다. 따라서 본 연구에서 개발된 KJC 예측기는 주어진 파라미터 값의 영역내에서 스스로 최적 파라미터 값을 추정해가면서 예측을 수행하도록 설계되었다. 자체 파라미터의 최적화는 외부 입력대신 모형의 연산과정에서 국부 최적화과정을 통해 수행된다. 이러한 국부 최적화 기반의 KNN-NPR 접근법은 기존 KNN-NPR기반 모형의 전역 최적화 문제를 해결함과 동시에 비교모형에 비하여 우수한 결과를 보였다.

KJC 예측기는 Off-line 적용 뿐만아니라 On-line 적용을 목적으로 개발되었다. 다시말해 자료의 수집에 발생하는 시간차짐을 모형에 고려하였기 때문에 구간 및 지점 검지체계에서 수집되는 모든 시계열자료를 적용할 수 있는 장점이 있다. 또한 신규 자료의 추가 및 입·출력 자료의 구조변경이 매우 용이하게 개발되었기 때문에 자료수집 환경과 모형의 적용 목적이 변화하는 경우 신속한 구조변경이 가능하다. 추가로 본 연구에서 개발된 모형은 복잡한 수학적 방정식 대신 이력자료의 검색과정과 군집화 과정을 통해 장래상태를 예측하게 됨으로 일반 ITS시스템의 자료를 유지·관리하는 운영요원들에게 친숙하다. 따라서 모형의 현장 적용성에 있어 타 예측모형에 비해 장점이 있다고 판단된다.

마지막으로 이력자료는 다량의 결측자료를 포함하고 있다. 따라서 이력자료를 이용하는 예측모형은 이력자료에 포함된 결측자료에 대한 처리를 모형 자체내에서 처리해야 하지만, 많은 연구를 이를 고려하지 않고 있다. 누락자료 처리기능이 결여된 모형을 On-line 예측에 적용하기 위해서는 누락자료의 보정(Imputation)을 위한 별도의 추정모형이 필요하다. 따라서 본 연구에서 개발된 KJC 예측기는 입력자료의 구축과 장래상태의 예측시 발생하는 결측자료의 문제를 모형 자체에서 처리함으로써 자료관리시스템에서 적용성을 극대화하였다고 판단된다.

이상에서 설명한 현장 적용성을 고려한 개발요소는 모듈화되어 유기적으로 결합되었다. 따라서 고려요소의 변경 및 추가가 용이하기 때문에 자료관리시스템에 탑재 및 운영시 다양한 자료환경의 변화 및 목적에 따라서 개발모형의 변경이 가능하다.

VI. 결론 및 향후 연구

1. 결론

스마트폰 보급대수 4,000만 대와 도시교통 대용량 자료의 시대는 지능형 교통체계(Intelligent transport system, 이하 ITS)의 예측 모델링 분야에서 변화와 기회를 강요하고 있다. 그러나 20세기말 ITS가 도입된 이후로 ITS 예측 모델링은 15분 이내를 예측하는 단기예측의 수준에 머물고 있으며, 그 이면에는 장래 상태의 불확실성 극복이라는 풀리지 않는 숙제가 있다. 본 연구는 강요된 변화와 기회에 대한 도전의 시작점이라 할 수 있다.

예측 모델링의 자료환경은 자료관리시스템에서 관리되는 대용량 자료 지원의 환경으로 변화하였다. 따라서 본 연구에서는 기존의 인위적인 예측 모델링 대신 방대한 양의 교통 데이터베이스에 내재된 지식을 이용하여 장래 교통정보를 생성하기 위한 KJC 예측기를 개발하였다. 개발된 예측기는 자료관리시스템 등에 탑재를 목표로 개발되었으며, 예측의 시·공간 목표는 스케줄 기반의 장래 교통정보와 장거리 경로통행시간 정보를 제공할 수 있도록 도로망 소통상태를 예측하는 것이다. KJC 예측기는 네트워크 소통상태 변수를 입력자료로 네트워크 소통상태 변수를 예측하는 다중 입·출력 구조와 강력한 패턴인식을 기반으로 개발되었다. 개발모형의 성능평가 결과 우수한 예측력을 보였으며, 본 연구의 결론과 시사점을 요약하면 다음과 같다.

첫째, 네트워크 소통상태로 이용된 통행의 방향성과 통행속도의 시·공간적 진화 상태는 장래 상태에 대한 충분한 기준선(Baseline)을 제공함으로써 불확실성을 감소시켰다. 장래 상태의 불확실성 감소는 높은 예측 정확도로 이어졌으며, 네트워크 소통상태를 최소 장래 6시간까지 예측할 수 있었다. 이는 기존 단기예측(Short-term forecasting)의 예측영역을 극복할 수 있는 가능성 뿐만 아니라 예측분야 연구에 있어 대용량 입력자료의 잠재력(Potentiality)에 대한 직·간접적인 증거이다.

둘째, 개발된 KJC 예측기는 혼돈 체계 이론을 근간으로 하고 강력한 패턴인식을 기능을 통해 장래 상태를 예측하는 KNN-NPR 접근법을 이용하여 개발되었다. 그리고 개발모형의 우수한 예측력은 상태벡터로 이용된 도로망에서 통행의 방향성과 통행속도의 상태가 초기 결정성(Initial deterministicity)이 강한 혼돈상태(Chaotic state)라는 간접적 증거이다. 또한 네트워크 소통상태는 패턴인식에 있어 효과적인 Attractor라고 판단된다.

셋째, 네트워크 소통상태를 다중입력체계로 이용하는 KJC 예측기의 국부 파라미터 최적화 전략은 장래 상태의 불확실성을 현저히 감소시키면서 우수한 예측력을 보였다. 이는 도로망 소통상태의 시·공간적 진화 특성이 분명하든 분명하지 않든 패턴의 경계가 존재한다는 간접적인 증거이다. 추가로 네트워크 소통상태는 단일 패턴의 진화가 아닌 패턴과 패턴이 연계되면서 진화하다는 사실을 간접적으로 확인하였다.

넷째, 진화상태를 이용한 패턴인식을 기반으로 하는 KJC 예측기는 도로망 혼잡의 생애주기(Life cycle)를 시·공간적으로 예측하였다. 이는 반복적이든 비 반복적이든 혼잡에 시·공간적 패턴이 존재함을 의미하며, 혼잡의 생애주기 패턴은 단일 또는 다수의 패턴이 중첩되면서 새로운 패턴을 형성하는 것으로 판단된다.

다섯째, 구간검지체계는 자료의 수집에 있어 필연적으로 시간처짐이 발생하며, 시간처짐의 길이는 단기 예측모형의 예측영역을 초과하는 경우가 발생한다. 개발된 KJC 예측기는 이러한 시간처짐을 고려한 On-line 예측기로 개발되었으며, 5~60분의 시간처짐을 효과적으로 극복하면서 장래상태를 예측하였다. 이는 단일 구간 또는 지점이라는 기존의 공간적 개념에서 벗어나 도로망의 개념으로 상태를 확장하였기 때문이다. 따라서 도로망 상태기반 입력자료 체계는 구간검지체계에서 타 예측접근법에 비하여 장점을 가진다고 판단된다.

여섯째, 예측영역의 확장은 시계열의 길이를 증가시킨다고 증가하지 않는다. 즉, 시계열 상태의 길이를 증가시킨다 해도 확장된 예측영역의

장래 상태에 대한 불확실성을 감소시킬 수는 없다. 이는 기존의 단일 지점 및 구간을 공간적 범위로 하는 예측기법이 단기예측의 예측영역을 확장할 수 없던 근본적인 이유라고 판단된다. 예측의 시간적 영역을 확장하기 위해서는 도로망에서 통행의 방향성과 소통상태와 같은 장래 상태에 대한 Indicator를 공간적으로 확장해야 한다.

일곱째, 본 연구에서 개발된 KJC 예측기는 예측의 시·공간 영역을 도로망의 장래 6시간까지 확장함으로써 물리학에 위배되지 않는 단·중·장거리 경로통행시간을 예측하였다. 예측의 정확도는 단·중·장거리 모든 부분에서 기존의 방법론보다 우수한 성능을 보였다. 따라서 통행시간과 거리에 한정되지 않는 경로통행시간 예측방법론이 개발되었다고 판단된다. 추가적으로 개발된 KJC 예측기는 강력한 패턴인식을 기반으로 하기 때문에 반복적 특성에 관계없이 다양한 형태의 경로통행시간 예측에 장점을 가진다고 판단된다.

마지막으로 KJC 예측기는 자료관리시스템의 탑재를 고려하여 연산수행속도를 극대화함과 더불어 모형의 구조변경, 입·출력자료의 구조변경의 용이성, 모형 자체의 결측자료 처리 기능, 파라미터 최소화 등 시스템의 적용성을 극대화하였다. 이러한 예측모형의 개발방향은 모형의 적용성 뿐만 아니라 기술적 발전에도 기여할 것으로 판단된다. 추가로 방대한 양의 자료가 수집·관리되고 있음에도 불구하고 이를 활용한 장래 상태 예측기술은 개발되지 않고 있는 실정이다. 따라서 본 연구에서 개발된 KJC 예측기는 이력자료의 활용성과 더불어 기존 교통정보 예측기술에 대한 패러다임 전환을 모색했다는 데 의의가 있다.

2. 향후 연구

KJC 예측기의 활용방안과 고도화 측면에서 지속적인 향후 연구가 필요하다. 활용방안으로는 첫째, 기존의 KNN-NPR관련 예측연구에 적용이 가능하다. 개발된 예측기는 강력한 패턴인식을 기반으로 국부 파라미터

최적화를 수행한다. 따라서 다중시간대 교통변수 예측, 다변량 교통변수 예측, 극심한 변동을 보이는 단속류 교통류 예측, 버스정류소간 운행소요 시간 예측 등에 직접적 적용이 가능하다. 둘째, 본 연구에서 예측의 시간적 범위는 경로통행시간의 예측을 위해 6시간까지 확장되었다. 그러나 KJC 예측기는 6시간 이후까지 예측이 가능하다. 또한 30분/1시간/2시간/4시간 이후의 스케줄 기반 경로통행시간 예측과 정보제공이 가능하다. 셋째, KJC 예측기는 도로망의 소통상태를 이용한 패턴인식에 기초하며 구간검지체계에 적합하도록 설계되었다. 따라서 반복적 행태가 강한 도시부의 장래 소통정보 예측 및 교통예보 등에 적극 활용할 수 있다. 마지막으로 KJC 예측기의 다중 입·출력 체계와 패턴인식 기능을 이용하여 누락자료의 보정(Imputation)에 직접적인 활용이 가능하다.

KJC 예측기의 고도화를 위한 향후 연구는 다음과 같다. 첫째, 예측된 도로망 소통상태는 현행자료와 융합되어 보다 정확한 예측이 가능하다. 따라서 자료융합기법 등을 이용하여 예측 정확도를 극대화할 필요가 있다. 둘째, KNN의 상태벡터는 규정되어 있지 않다. 따라서 상태 재구성(Reconstruction)을 위한 무한한 상태벡터가 존재하며, 이를 통한 예측성능의 향상에 대한 기회가 충분하다. 셋째, 입력자료의 공간적 영역인 네트워크의 크기 및 도로구간별 가중치를 이용한 고도화가 필요하다. 이는 패턴인식과 깊이 관련되어 있기 때문에 KJC 예측기의 예측성능을 향상시킬 수 있는 많은 여지가 있다. 추가로 정보 필터링 기능 등을 이용하여 상태인식을 향상시킬 수 있다. 넷째, KNN은 다양한 예측식을 적용할 수 있다. 본 연구에서는 상태간 거리의 역수를 이용하였으나 또 다른 가능성이 존재한다. 다섯째, 상태간 유사성(Similarity)과 인접성(Nearness)을 구분하는 상태간 거리 및 통계적 기법을 이용한 패턴인식의 기능 향상의 여지가 있다. 마지막으로 다중 입출력 체계를 기반으로 개발되었기 때문에 이중자료와의 결합이 용이하며 이에 대한 지속적인 연구가 필요하다.

참고문헌

[국내연구]

- 강정규, 남궁성(2002), “고속도로 통행료수납자료를 이용한 통행시간 예측모형 개발”, 대한교통학회지, 20권 4호, pp. 151-162.
- 강진기, 손영태, 윤여환, 변상철(2002), “비매설식 자동차량인식장치를 이용한 구간교통정보 산출 방법 연구”, 한국ITS학회논문지, 1권 1호, pp. 22-32.
- 김재진, 노정현, 남궁성, 박동주(2007), “출발시각기준 링크통행시간 정보의 공간적 설계(연속류를 중심으로)”, 대한교통학회지, 25권 2호, pp. 145-155.
- 남궁성, 노정현(1996), “전문가시스템을 이용한 최적경로 탐색시스템(X-PATH)의 개발”, 대한교통학회지, 14권 1호, pp. 51-67.
- 남궁성(2005), “고속도로 경로통행시간 산출을 위한 전진반복전후방탐색법(PIFAB)의 개발”, 대한교통학회지, 23권 5호, pp. 147-155.
- 박동주(1999), “통행시간 산정 및 예측을 위한 최적 집계시간간격 결정에 관한 연구”, 1999년도 제36회 학술발표회논문집, pp. 440-445.
- 심상우, 최기주, 이상수, 남궁성(2013), “하이패스 DSRC 자료를 활용한 고속도로 오프라인 경로통행시간 추정기법 개발”, 대한교통학회지, 31권 3호, pp. 45-54.
- 이승재, 장현호(2003), “CA모형을 이용한 단기 구간통행시간 예측에 관한 연구”, 대한교통학회지, 21권 1호, pp. 91-102.
- 이영인, 이정희(2002), “교통정보제공을 위한 구간통행시간 산출 방법론 연구 (적정표본수를 중심으로)”, 대한교통학회지, 20권 3호, pp. 55-67.
- 장현호(2002), 『Cellular Automata 모형을 이용한 대규모 네트워크 단기링크통행시간 예측에 관한 연구』, 서울시립대학교 석사학위논문.
- 장현호, 백승걸, 박재범(2004), “다양한 연속 교통류 구현을 위한 확률과장전과모형의 개발” 대한교통학회지, 22권 4호, pp. 147-158.
- 장현호, 한동희, 이태경, 이영인, 원제무(2010), “NPR기반 누락 교통자료 추정기법 개발 및 적용” 대한교통학회지, 28권 3호, pp. 61-74.

[국외연구]

- Abdulhai, B., Porwal, H., and W. Recker(1999), “Short term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks”, Transportation Research Board Annual Meeting.
- Abu-Lebdeha, G., and A.K. Singh(2011), “Modeling arterial travel time with limited traffic variables using conditional independence graphs & state-space neural networks”, Procedia Social and Behavioral Sciences Vol. 16, pp. 207 - 217.
- Adeli, H.(2001), “Neural networks in civil engineerin”, Computer-Aided Civil and Infrastructure Engineering, Vol. 16(2), pp. 126-142.
- Altman, N.S.(1992), “An introduction to kernel and nearest-neighbor nonparametric regression”, The American Statistician, Vol. 46, pp. 175-185.
- Amin, S.M., Rodin, E.Y., Liu, A.P., and K. Rink(1998), “Traffic prediction and management via RBF neural nets and semantic control”, Computer -Aided Civil and Infrastructure Engineering, Vol. 13, pp. 315-327.
- Appert-Rolland, C., Chevoir, F., Gondret, P., Lassarre, S., Lebacque, J-P, and M.S. Schreckenberg(2008), Traffic and Granular Flow '07, Springer.
- An, S., Chang, H., and Y. Lee(2013), “Dynamic forecasting of bus path travel time: simplified stochastic approach”, International Journal of Information Technology and Business Management, Vol. 19, pp. 28-34.
- Arthur, D., and S. Vassilvitskii(2007), “ k -means++: The advantages of careful seeding”, in Proceeding 18th Annual ACM-SIAM Sympion Discrete Algorithms, Philadelphia, PA, pp. 1027-1035.
- Ball, G.H. and D.J. Hall(1965), ISODATA, A novel technique for data

analysis, and pattern classification, Technical Report, Menlo Park, California: Stanford Research Institute.

- Ben-Akiva, M., Bierlaire, M., and D. Burton(2001), “Network state estimation and prediction for real-time traffic management”, *Networks and Spatial Economics*, Vol. 1, pp. 293–318.
- Bigazzi, A.Y., Siri, H., and R.L. Bertini(2010), “Effects of temporal data aggregation on performance measures and other ITS applications”, presented at the 89th Annual Meeting of Transportation Research Board, Washington, DC.
- Box, G.E.P. and G.H. Jenkins(1976), *Time Series Analysis Forecasting and Control*, Revised Edition, San Francisco, Holden Day.
- Bustillos, B.I., and Y.C. Chiu(2011), “Real-Time Freeway-Experienced Travel Time Prediction Using N-Curve and k Nearest Neighbor Methods”, *Transportation Research Record*, Vol. 2243, pp. 127–137.
- Caliliski, T. and Harabasz, J.(1974), “A dendrite method for cluster analysis”, *Communications in Statistics*, Vol. 3(1), pp. 1–27.
- Chang, H., Baek, S., Namkoong, S.J., and B. Yoon(2005), “Some findings of CA model to generate various freeway traffic flows with additional rules, *Journal of EASTS*, Vol. 6, pp. 1368–1381.
- Chang, H., Baek, S., Kim, H., Shah, A.A., Lee, J.D., and N. P. Mahalik (2007), “Development of distributed real-time decision support system for traffic management centers using microscopic CA model”, *Iranian Journal of Science & Technology, Transaction B, Engineering*, Vol. 31(B2), pp. 155–166.
- Chang, H., Park, D., Lee, S., Lee, H., and S. Baek(2010), “ Dynamic

multi-interval bus travel time prediction using bus transit data”, *Transportmetrica*, Vol. 6(1), pp. 19–38.

- Chang, H., Seong, J.N., Lee, Y., and B. Yoon(2011), “Dynamic freeway path travel time prediction based on nonparametric regression approach using dedicated short-range communications data”, presented at the 90th Annual Meeting of Transportation Research Board, Washington, DC.
- Chang, H., Park, D., Lee, Y., and B. Yoon(2012_a), “Multiple time period imputation technique for multiple missing traffic variables: non parametric regression approach”, *Canadian Journal of Civil Engineering*, Vol. 39, pp. 448–459.
- Chang, H., Lee, Y., Yoon, B., and S. Baek(2012_b), “Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences”, *IET: Intelligent Transport Systems*, Vol. 6(3), pp. 292–305.
- Chen, M., and S. Chein(2000), “Determining the number of probe vehicles for freeway travel estimation using microscopic simulation”, *Transportation Research Record*, No. 1719, pp. 61–68.
- Chen, H., and S. Grant-Muller(2001), “Use of sequential learning for short-term traffic flow forecasting”, *Transportation Research Part C*, Vol. 9, pp. 319–336.
- Chien, S., Ding, Y., and C. Wei(2002), “Dynamic bus arrival time prediction with artificial neural network”, *ASCE Journal of Transportation Engineering*, Vol. 128, pp. 429–438.
- Chien, S.I.J., and C.M. Kuchipudi(2003), “Dynamic travel time prediction with real-time and historic data”, *Journal of transportation engineering*, Vol. 129(6), pp. 608–616.

- Chrobok, R., Kaumann, O., Wahle, J., and M. Schreckenberg(2004_a), “Different methods of traffic forecast based on real data”, European Journal of Operational Research, Vol. 155, pp. 558–568.
- Chrobok, R., and Pottmeier, A., and S.F. Hafstein(2004_b), “Traffic forecast in large scale freeway networks”, International Journal of Bifurcation and Chaos, Vol. 14(6), pp. 1995–2004.
- Clark, S.(2003), “Traffic prediction using multivariate nonparametric regression”, Journal of Transportation Engineering, Vol. 129(2), pp. 161–168.
- Coifman, B., Dhoorjaty, S., and Z.H. Lee(2003), “Estimating median velocity instead of mean velocity at single loop detectors”, Transportation Research Part C, Vol. 11, pp. 211–222.
- Coufal, D., and E. Turunen(2004), “Short term prediction of highway travel time using data mining and neuro-fuzzy methods”, Neural Network World, Vol. 3(4), pp. 221–231.
- Cremer, M.(1995), “On the calculation of individual travel times by macroscopic models”, in the Proceedings of the 6th International VNIS Conference.
- Dailey, D., and N. Taiyab(2002), A cellular automata model for use with real freeway data, TRAC/WSDOT Final Technical Report WA-RD 537.1., Washington State Transportation Center.
- Davis, G. and N. Nihan(1991), “Nonparametric regression and short-term freeway traffic forecasting”, Journal of Transportation Engineering, Vol. 117, pp. 178–188.
- Devijver, P.(1982), Statistical pattern recognition. Applications of pattern recognition, K. S. Fu, ed., CRC Press, Boca Raton, Fla., pp. 15–36.

- Dion, F., and H. Rakha(2003), “Estimation spacial travel time using automatic vehicle identification data”, presented in the 82nd Annual Meeting of Transportation Research Board, Washington, DC.
- Dion, F., and H. Rakha(2006), “Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates”, *Transportation Research Part B*, Vol. 40(9), pp. 745 - 766.
- Disbro, J.E. and M. Frame(1989), *Traffic flow theory and chaotic behavior*, New York State Department of Transportation Report FHWA/NY/SR-98/91, New York.
- Eubank, J.D.(1988), *Spline smoothing and nonparametric regression*, Marcel Dekker Inc, NY.
- Farmer, J.D., and J.J. Sidorowich(1987), “Predicting chaotic time series”, *Physical Review Letter*, Vol. 59, pp. 845-848.
- Fei, X., Lu, C.C., and K.A. Liu(2011), “Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction”, *Transportation Research Part C*, Vol, 19(6), pp. 1306-1318.
- Fukui, M., Sugiyama, Y., Schreckenberg, M., and D.E. Wolf(2002), *Traffic and Granular Flow '01*, Springer.
- Guegan, D., and J. Leroux(2009), “Forecasting chaotic systems: The role of local Lyapunov exponents”, *Chaos, Solitons & Fractals*, Vol. 41, pp. 2401-2404.
- Guo, J.(2005), *Adaptive estimation and prediction of univariate vehicular traffic condition series*, Ph.D. Dissertation, North Carolina State University, Civil Engineering.
- Hamad, K., Lee, E., Shourijeh, M.T., and A. Faghri(2009), “Near-term

travel speed prediction utilizing Hibert-Huang transform”, Computer -Aided Civil and Infrastructure Engineering, Vol. 24, pp. 551-576.

- Hartigan, J.A.(1975), Clustering Algorithms, John Wiley & Sons, New York, London, Sidney, Toronto.
- Haworth, J., and T. Cheng(2012), “Non-parametric regression for space -time forecasting under missing data”, Computers, Environment and Urban Systems, Vol. 36(6), pp. 538 - 550.
- Hoogendoorn, S.P., Luding, S., Bovy, P.H., Schreckenberg, M., and D.E. Wolf(2004), Traffic and Granular Flow '03, Springer.
- Hounsell, N.B., and S. Ishtiaq(1997), “Journey time forecasting for dynamic route guidance systems in incident conditions”, International Journal of Forecasting, Vol. 13 pp. 33-42.
- Huang, N., Zheng, S., Long, S., Wu, M., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., and H. Liu(1998), “The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis”, Proceedings of the Royal Society of London: Series A, Mathematical and Physical Sciences, Vol. 454, pp. 93-95.
- Huiskens, G.(2003), “Soft-computing techniques applied to short-term traffic flow forecasting”, Systems analysis modelling simulation, Vol. 43(2), pp. 165-173.
- Innamaa, S.(2000), “Short-term prediction of traffic situation using MLP-neural networks”, Proceeding in 7th World Congress of Intelligent Transportation System, Turin, Italy.
- Innamaa, S.(2005), “Short-term prediction of travel time using neural networks on an interurban highway”, Transportation, Vol. 32, pp. 649-669.

- Innamaa, S.(2006), “Effect of monitoring system structure on short-term prediction of highway travel time”, *Transportation Planning and Technology*, Vol. 29(2), pp. 125–140.
- Innamaa, S.(2009), “Self-adapting traffic flow status forecasts using clustering”, *IET Intelligent Transport Systems*, Vol. 3(1), pp. 7–76.
- Ishak, S., and C. Alecsandru(2004), “Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings”, *Journal of Transportation Engineering*, Vol. 130(4), pp. 452–465.
- Jiang, X., and H. Adeli(2005), “Dynamic wavelet neural network model for traffic flow forecasting”, *Journal of transportation engineering*, Vol. 131(10), pp. 771–779.
- Jintanakul, K., Chu, L., and R. Jayakrishnan(2009), “Bayesian mixture model for estimating freeway travel time distributions from small probe samples from multiple days”, *Transportation Research Record*, Vol. 2136, pp. 37–44.
- Juri, N.R., Unnikrishnan, A., and S.T. Waller(2007), “Integrated traffic simulation–statistical analysis framework for online prediction of freeway travel time”, *Transportation Research Record*, Vol. 2039, pp. 24–31.
- Kalman, R.E.(1960), “A new approach to linear filtering and prediction problems”, *Journal of Basic Engineering, Transactions of the ASME* Vol. 82D, pp. 35–45.
- Kalman, R.E., and R.S. Bucy(1961), “New results in linear filtering and prediction theory”, *Journal of Basic Engineering*, Vol. 60, pp. 95–108.
- Kamarianakis, Y. and P. Prastacos(2004), “Forecasting traffic flow conditions

in an urban network: comparison of multivariate and univariate approaches”, *Transportation Research Record*, Vol. 1857, pp. 74–84.

- Karlaftis, M.G., and E.I. Vlahogianni(2011), “Statistical methods versus neural networks in transportation research: Differences, similarities and some insights”, *Transportation Research Part C*, Vol. 19, pp. 387–399.
- Karlsson, M., and S. Yakowitz(1987), “Rainfall-runoff forecasting methods, old and new”, *Stochastic Hydrology and Hydraulics*, Vol. 1, pp. 303–318.
- Kaufman, L., and P.J. Rousseeuw(1990), *Finding groups in data: An introduction to cluster analysis*, Wiley, New York.
- Kennel, M.B., Brown, R., and H.D. Abarbanel(1992), “Determining embedding dimension for phase-space reconstruction using a geometrical construction”, *Physical Review A*, Vol. 45(6), pp. 3403–3411.
- Kerner, B.S.(2002), “Three-Phase Traffic Theory”, *Traffic and Granular Flow '01*, pp. 13–50.
- Kerner, B.S, and H. Rehborn(1996), “Experimental properties of complexity in traffic flow”, *Physical Review E*, Vol. 53(5), pp. 4276–4278.
- Khan, A.M.(2012), “Bayesian predictive travel time methodology for advanced traveler information system”, *Journal of Advanced Transportation*, Vol. 46(1), pp. 67–79.
- Khosravi, A., Mazloumi, E., Nahavandi, S., Creighton, D., and J.W.C. van Lint(2011), “Prediction intervals to account for uncertainties in travel time prediction”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12(2), pp. 537–547.
- Kim, D., Park, D., Rho, J., Baek, S., and S. Namkoong(2007), “A study on the construction of past travel time pattern for freeway travel time

forecasting: focused on loop detectors”, *International Journal of Urban Sciences*, Vol. 11(1), pp. 14-29.

- Kim, H., Akhtar Ali Shah, S., Chang, H., and H. Ahn(2005), “Microscopic simulation in decision support system for the incident induced traffic management”, *Knowledge-Based Intelligent Information and Engineering Systems(LNCS/LNCI)*. LNAI 3681, pp. 255-260.
- Kim, J., Rho, J., and D. Park(2009), “On-line estimation of departure time-based link travel times from spatial detection system”, *International Journal of Urban Sciences*, Vol. 13(1), pp. 63-80.
- Kim, M., Kho, S., and C. Park(2011), “Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system”, *Transportation Research Record*, 2256, pp. 51-59.
- Kirby, H.R., Watson, S.M., and M.S. Dougherty(1997), “Should we use neural networks or statistical models for short-term motorway traffic forecasting?”, *International Journal of Forecasting*, Vol. 13, pp. 43-50.
- Kuchipudi, C.M., and S.I.J. Chien(2002), “Development of a hybrid model for dynamic travel-time prediction”, *Transportation Research Record*, Vol. 1855, pp. 22-31.
- Kwon, J., Coifman, B., and P. Bickel(2000), “Day-to-day travel time trends and travel time prediction from loop detector data”, *Transportation Research Record*, Vol. 1717, pp. 120-129.
- Kwon, J, and K. Petty(2005), “A Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes”, *Transportation Research Record*, Vol. 1935, pp. 147-153.

- Lam, W.H.K., Chan, K.S., and J.W.Z. Shi(2002), “A traffic flow simulator for short-term travel time forecasting”, *Journal of Advanced Transportation*, Vol. 36(3), pp. 265–291.
- Lam, W.H.K., Tang, Y.F., and M.L. Tam(2006), “Comparison of two non-parametric models for daily traffic forecasting in Hong Kong”, *International Journal of Forecasting*, Vol. 25, pp. 173–192.
- Lan, C.J., and S.P. Miaou(1999), “Real-time prediction on traffic flows using dynamic generalized linear models”, *Transportation Research Record*, Vol. 1678, pp. 168–178.
- Lee, Y. and C. Choi(1998), “Development of link travel time prediction algorithm for urban expressway”, *Proceeding of 5th World Congress on ITS*, Korea.
- Lee, S., and D.B. Fambro(1999), “Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting”, *Transportation Research Record*, Vol. 1678, pp. 179–188.
- Li, Y.(2008), “Short-term prediction of motorway travel time using ANPR and loop data”, *Journal of Forecasting*, Vol. 27(6), pp. 507–517.
- Li, C.S., and M.C. Chen(2013), “Identifying important variables for predicting travel time of freeway with non-recurrent congestion with neural networks”, *Neural Computing & Applications*, Vol. 23(6), pp. 1611–1629.
- Li, L., Lin, W.H., and H. Liu(2006), “Type-2 fuzzy logic approach for short-term traffic forecasting”, *Intelligent Transport Systems*, IEE Proceedings.
- Li, R., and G. Rose(2011), “Incorporating uncertainty into short-term travel time predictions”, *Transportation Research Part C: Emerging*

Technologies, Vol. 19(6), pp. 1006–1018.

- Lin, W.H., Kulkarni, A., and P. Mirchandani(2004), “Short-term arterial travel time prediction for advanced traveler information systems”, *Intelligent Transportation Systems*, Vol. 8, pp. 143–154.
- Lingras, P., Sharma, S., and M. Zhong(2002), “Prediction of recreational travel using genetically designed regression and time-delay neural network models”, *Transportation Research Record*, Vol. 1805, pp. 16–24.
- Liu, Z., Sharma, S., and S. Datla(2008_a), “Imputation of missing traffic data during holiday periods. *Transportation Planning and Technology*”, Vol. 31(5), pp. 525–544.
- Liu, M, Yu, L., Geng, Y., and X. Chen(2008_b), “Double-sided optimization of ITS data aggregation via wavelet transformation”, *Journal of Transportation systems Engineering and information technology*, Vol. 8(1), pp. 49–54.
- Lloyd, S.P.(1982), “Least squares quantization in PCM”, *IEEE Transactions Information Theory*, Vol. 28, pp. 129–137.
- Lu, C.C.J.(2012), “An adaptive system for predicting freeway travel times”, *International Journal of Information Technology and Decision Making*, Vol. 11(4), pp. 727–747.
- Ma, Y., Chowdhury, M., Sadek, A., and M. Jaihani(2012), “Integrated traffic and communication performance evaluation of an intelligent vehicle infrastructure integration (VII) system for online travel-time prediction”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13(3), pp. 1369–1382.
- Marriott, F.H.C.(1971), “Practical problems in a method of cluster

analysis”, *Biometrics* 27, pp. 501–514.

- MacQueen, J.B.(1967), “Some methods for classification and analysis of multivariate observations”, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, University of California Press, Berkeley, Vol. 1, pp. 281–297.
- Mouskos, K.C., Niver, E., Pignataro, L.J., Lee, S., and N. Antoniou(1998), *TRANSMIT system evaluation. Final report*, Institute for Transportation, New Jersey Institute of Technology, N.J.
- Mu, T., Jiang, J., and Y. Wang(2012), “Heterogeneous delay embedding for travel time and energy cost prediction via regression analysis”, *Intelligent Transportation Systems, IEEE Transactions on Intelligent Transportation Systems*, Vol. 14(1), pp. 214–224.
- Mulhern, F.J., and R.J. Caprara(1994), “A nearest neighbor model for forecasting market response”, *International Journal of Forecasting*, Vol. 10, pp. 191–207.
- Nagel, K., and M. Schreckenberg(1992), “A cellular automaton model for freeway traffic”, *Journal de Physique I*, Vol. 2, pp. 2221–2229.
- Nicholson, H., and C.D. Swann(1974), “The prediction of traffic flow volumes based on spectral analysis”, *Transportation Research*, Vol. 8, pp. 533–538.
- Oh, C., and S. Park(2011), “Investigating the effects of daily travel time patterns on short-term prediction”, *KSCE Journal of Civil Engineering*, Vol. 15(7), pp. 1263–1272.
- Oh, C., Ritchie, S.G., and J.S. Oh(2005), “Exploring the relationship between data aggregation and predictability toward providing better

predictive traffic information”, Transportation Research Record, Vol. 1935, pp. 28–36.

- Ohba, Y., Ueno, H., and M. Kuwahara, “Travel time calculation method for expressway using toll collection system data”, Proceedings in IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, Tokyo, Japan.
- Okutani, I., and Y.J. Stephanedes(1984), “Dynamic prediction of traffic volume through Kalman filtering theory”, Transportation Research Part B, Vol. 18B(1), pp. 1–11.
- Oliver, D., and E. Nour-Eddin(2006), Innovative Processing of Toll Collection Data, LICIT report, No. 0604.
- Oswald, R.K., Scherer, W.T., and B. Smith(2000), Traffic flow forecasting using approximate Nearest Neighbor Nonparametric regression, A research project report for U.S. DOT, University transportation center.
- Packard, N.H., Crutchfield, J.P., Farmer, J.D., and R.S. Shaw(1980), “Geometry from a time series”, Physical Review Letters, Vol. 45, pp. 712–716
- Park, B., Messer, C.J., and T. Urbanik II(1998), “Short-term freeway traffic volume forecasting using radial basis function neural networks”, Transportation Research Record, Vol. 1651, pp. 39–47.
- Park, D., and L.R. Rilett(1998), “Forecasting multiple-period freeway link travel times using modular neural networks”, Transportation Research Record, Vol. 1617, pp. 163–170.
- Park, D., and L.R. Rilett(1999), “Forecasting freeway link travel times with a multilayer feedforward neural network”, Computer-Aided Civil

and Infrastructure Engineering, Vol. 14, pp. 357–367.

- Park, D., and Rilett, L.R., and G. Han(1999), “Spectral basis neural network for real-time travel time forecasting”, Journal of Transportation Engineering, Vol. 125(6), pp. 515–523.
- Peeta, S., and H.S. Mahmassani(1995), “System optimal and user equilibrium time-dependent traffic assignment in congested networks”, Annals of Operation Research, Vol. 60, pp. 81–113.
- Qi, Y., and B.L. Smith(1987), “Identifying nearest-neighbors in large-scale incident data archive”, Transportation Research Record, Vol. 1987, pp. 89–98.
- Qiao, W., Haghani, A., and M. Hamed(2012), “Short-Term Travel Time Prediction Considering the Effects of Weather”, Transportation Research Record: Vol. 2308(1), pp. 61–72.
- Ran, B. Boyce, D.E., and L.J. LeBlanc(1993), “A new class of instantaneous dynamic user-optimal traffic assignment models”, Operation Research, Vol. 41(1), pp. 192–202.
- Rice, J., and E. van Zwet(2004), “A simple and effective method for predicting travel times on freeways”, IEEE Transactions on Intelligent Transportation Systems, Vol. 5(3), pp. 200–207.
- Rilett, L.R., and D. Park(2001), “Direct forecasting of freeway corridor travel times using spectral basis neural networks”, Transportation Research Record, Vol. 1752, pp. 140–147.
- Robinson, P.(1983), “Nonparametric estimators for time series”, Journal of Time Series Analysis, Vol. 4, pp. 185–207.
- Schadschneider, A., Poschel, T., Kuhne, R., Schreckenberg, M., and D.E. Wolf(2006), Traffic and Granular Flow ‘05, Springer.

- Shah, A.A., Kim, H., Baek, S. Chang, H., and B. Ahn(2008), “System architecture of a decision support system for freeway incident management in Republic of Korea”, *Transportation Research Part A*, Vol. 42(5), pp. 799-810.
- Simroth, A., and H. Zahle(2011), “Travel time prediction using floating car data applied to logistics planning”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12(1), pp. 243-253.
- Smith, B.L.(1995), “Forecasting freeway traffic flow for intelligent transportation system applications”, Doctoral dissertation, Department of Civil Engineering, University of Virginia, Charlottesville, VA.
- Smith, B.L., and J.H. Conklin(2002), “Use of local lane distribution patterns to estimate missing data values from traffic monitoring systems”, *Transportation Research Record*, Vol. 1811, pp. 50 - 56.
- Smith, B.L., and M.J. Demetsky(1995), “Short-term traffic flow prediction: neural network approaches”, *Transportation Research Record*, Vol. 1453, pp. 98-104.
- Smith, B.L., and M.J. Demetsky(1996), “Multiple-interval freeway traffic flow forecasting”, *Transportation Research Record*, Vol. 1554, pp. 136-141.
- Smith, B.L., and M.J. Demetsky(1997), “Traffic flow forecasting: comparison of modeling approaches”, *Journal of Transportation Engineering*, Vol. 123(4), pp. 261-266.
- Smith, B.L., and R.K. Oswald(2003), “Meeting real time traffic flow forecasting requirements with imprecise computations”, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 18(13), pp. 201-213.
- Smith, B.L., Scherer, W.T., and J.H. Conklin(2003), “Exploring im-

putation techniques for missing data in transportation management systems”, *Transportation Research Record*, Vol. 1836(1), pp. 132 - 142.

- Smith, B.L., and J.M. Ulmer(2003), “Freeway traffic flow rate measurement: Investigation into impact of measurement time interval”, *Journal of Transportation Engineering*, Vol. 129(3), pp. 223-229.
- Smith, B.L., Williams, B.M., and R.K. Oswald(1999), “Parametric and nonparametric traffic volume forecasting”, *Proceeding in the 78th Annual Meeting of Transportation Research Board*, Washington, DC.
- Smith, B.L., Williams, B.M., and R.K. Oswald(2002), “Comparison of parametric and nonparametric models for traffic flow forecasting”, *Transportation Research Part C*, Vol. 10, pp. 303-321.
- Soriguera, F., and F. Robuste(2011), “Highway travel time accurate measurement and short-term prediction using multiple data sources”, *Transportmetrica*, Vol. 7(1), pp. 85-109.
- Soriguera, F., Thorson, L., and F. Robuste(2007), “Travel time measurement using toll infrastructure”, *Transportation Research Record* 1981, pp. 99-107.
- Stathopoulos, A., Dimitriou, L., and T. Tsekeris(2008), “Fuzzy modeling approach for combined forecasting of urban traffic flow”, *Computer-Aided Civil Infrastructure Engineering*, Vol. 23(7), pp. 521 - 535.
- Stathopoulos, A., and M.G. Karlaftis(2001), “Temporal and spatial variations of real-time traffic data in urban areas”, *Transportation Research Record*, Vol. 1768, pp. 135-140.
- Sun, H., Liu, X., Xiao, H., He, R.R., and B. Ran(2003), “Use of local linear regression model for short-term traffic forecasting”, *Transportation*

Research Record, Vol. 1936, pp. 143–150.

- Takens, F.(1981), “Detecting strange attractors in turbulence”, Dynamical Systems and Turbulence, Vol. 898, pp. 366–381.
- Thorndike, R.L.(1953), “Who belongs in the family?”, Psychometrika, Vol. 18(4), pp. 268–276.
- Tong, H.(1993), Nonlinear time series: a dynamic approach, Oxford University Press, New York.
- Turochy, R.E.(2006), “Enhancing short-term traffic forecasting with traffic condition information”, Journal of Transportation Engineering, Vol. 132(6), pp. 469–474.
- van Aerde, M.(1985), Modeling of traffic flows, assignment and queuing in integrated freeway/traffic signal networks, Ph.D. Thesis, Department of Civil Engineering, University of Waterloo, Waterloo, Canada.
- van Hinsbergen, C.P.I.J, van Lint, J.W.C., and F.M. Sanders(2007), “Short term traffic prediction models”, presented in ITS World Congress, Beijing, China.
- van Hinsbergen, C.P.I.J., van Lint, J.W.C., and H.J. van Zuylen(2009), “Bayesian training and committees of state-space neural networks for online travel time prediction”, Transportation Research Record Vol. 2105, pp. 118–126.
- van Lint, J.W.C.(2006), “Reliable real-time framework for short-term freeway travel time prediction”, Journal of Transportation Engineering, Vol. 132(12), pp. 921–932.
- van Lint, J.W.C.(2008), “Online learning solutions for freeway travel time prediction”, IEEE Transactions on Intelligent Transportation Systems,

Vol. 9(1), pp. 38–47.

- van Lint, J.W.C, Hoogendoorn, S.P., and H.J. van Zuylen(2002), “Freeway travel time prediction with state-space neural networks”, *Transportation Research Record*, Vol. 1811, pp. 30–39.
- van Lint, J.W.C., Hoogendoorn, S.P., and H.J. van Zuylen(2005), “Accurate travel time prediction with state-space neural networks under missing data”, *Transportation Research Part C: Emerging Technology*, Vol. 13(5/6), pp. 347–369.
- van Lint, J.W.C., and H.J. van Zuylen(2005), “Monitoring and predicting freeway travel time reliability”, *Transportation Research Record*, Vol. 1917, pp. 54–62.
- Vlahogianni, E.I., Golias, J.C., and M.G. Karlaftis(2004), “ Short-term traffic forecasting: overview of objectives and methods”, *Transport Reviews*, Vol. 24(5), pp. 533–557.
- Vlahogianni, E.I., Karlaftis, M.G., and J.C. Golias(2005), “ Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach”, *Transportation Research Part C*, Vol. 13(3), pp. 211–234.
- Vlahogianni, E.I., Karlaftis, M.G., and J.C. Golias(2006), “Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume”, *Transportation Research Part C*, Vol. 14(5), pp. 351–367.
- Vlahogianni, E.I., Karlaftis, M.G., and J.C. Golias(2014), “Short-term traffic forecasting: where we are and where we’re going”, *Transportation Research Part C*, Vol. 43, pp. 3–19.
- Voort, M.v.d., Dougherty, M., and S. Watson(1996), “Combining Kohonen

maps with ARIMA time series models to forecast traffic flow”, *Transportation Research*, Vol. 4(5), pp. 307–318.

- Wang, Z., and C. Liu(2005), “An empirical evaluation of the loop detector method for travel time delay estimation”, *Journal of Transportation Systems: Technology, Planning, and Operations*, Vol. 9(4), pp. 161–174.
- Wang, Y., Papageorgiou, M., and A. Messmer(2006), “RENAISSANCE - A unified macroscopic model-based approach to real-time freeway network traffic surveillance”, *Transportation Research Part C: Emerging Technologies*, Vol. 14(3), pp. 190–212.
- Wang, J., and Q. Shi(2012), “Short-term traffic speed forecasting hybrid model based on Chaos-Wavelet analysis-support vector machine theory”, *Transportation Research Part C: Emerging Technologies*, Vol. 27, pp. 219–232.
- Whittaker, J., Garside, S., and K. Lindveld(1997), “Tracking and predicting a network traffic process”, *International Journal of Forecasting*, Vol. 13, pp. 51–61.
- Williams, B.M.(1999), “Multivariate vehicular traffic flow prediction”, *Transportation Research Record*, Vol. 1776, pp. 194–200.
- Williams, B.M., Durvasula, P.K., and D.E. Brown(1998), “Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models”, *Transportation Research Record*, Vol. 1644, pp. 132–141.
- William, B.M., and L.A. Hoel(2003), “Modelling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results”, *Journal of Transportation Engineering*, Vol. 129(6), pp. 664–672.
- Wu, C.H., Ho, J.M., and D.T. Lee(2004), “Travel-time prediction with support

vector regression”, IEEE Transactions on Intelligent Transportation Systems, Vol. 5(4), pp. 276–281.

- Xia, J., Chen, M., and W. Huang(2011), “A multistep corridor travel-time prediction method using presence-type vehicle detector data”, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Vol. 15(2), pp. 104–113.
- Xie, Y., and Y. Zhang(2006), “A wavelet network model for short-term traffic volume forecasting”, Journal of Intelligent Transportation Systems, Vol. 10(3), pp. 141–150.
- Xu, L.(1997), “Bayesian Ying-Yang Machine, Clustering and Number of Clusters”, Pattern Recognition Letters, Vol. 18(11–13), pp. 1167–1178.
- Yakowitz, S.(1987), “Nearest-neighbor methods for time-series analysis”, Journal of Time Series Analysis, Vol. 8(2), pp. 235–247.
- Yang, J.S.(2005), “Travel time prediction using the GPS test vehicle and Kalman filtering techniques”, Proceedings of the 2005 American Control Conference.
- Yang, M., Liu, Y., and Z. You(2010), “The reliability of travel time forecasting”, IEEE Transactions on Intelligent Transportation Systems, Vol. 11(1), pp. 162–171.
- Yin, H., Wong, S.C., Xu, J., and C.K. Wong(2002), “Urban traffic flow prediction using a fuzzy-neural approach”, Transportation Research Part C, Vol. 10, pp. 85–98.
- Yoon, B. and H. Chang(2014), “Potentialities of data-driven non-parametric regression in urban signalized traffic flow forecasting”, Journal of Transportation Engineering, Vol. 140(7), pp. 1–13.

- You, J., and T.J. Kim(2000), “Development and evaluation of a hybrid travel time forecasting model”, *Transportation Research Part C*, Vol. 8, pp. 231–256.
- Zhang, X., and J.A. Rice(2003), “Short-term travel time prediction”, *Transportation Research Part C*, Vol. 11, pp. 187–210.
- Zhang, H., Ritchie, S.G., and Z.P. Lo(2000), “Macroscopic modeling of freeway traffic using an artificial neural network”, *Transportation Research Record*, Vol. 1588, pp. 110–119.
- Zheng, W., Lee, D.H., and Q. Shi(2006), “Short-term freeway traffic flow prediction: Bayesian combined neural network approach”, *Journal of transportation engineering*, Vol. 132(2), pp. 114–121.
- Zheng, F., and H. van Zuylen(2012), “Urban link travel time estimation based on sparse probe vehicle data”, *Transportation Research Part C: Emerging Technologies*, Vol. 31, pp. 145–157.
- Zhong, M., Sharma, S., and P. Lingras(2005), “Refining genetically designed models for improved traffic prediction on rural roads”, *Transportation Planning and Technology*, Vol. 28(3), pp. 213–236.
- Zou, N., Wang, J., Chang, G.L., and J. Paracha(2009), “Field test of a travel-time prediction system with widely spaced detectors”, *Transportation Research Record*, Vol. 2129(1), pp. 62–72.

Abstract

Network-based Intercity Path Travel Time Forecasting Using Large-scale Data

Chang, Hyun-Ho

Department of Environmental Planning
The Graduate School of Environmental Studies
Seoul National University

Today's urban and transportation information can be summarized simply with the words 'big data'. Unimaginably tremendous quantities of information collected by smart-phones and information devices of 21C intelligent transportation systems (ITS) based on edge technology accounts for much of this big data, and changes in the data environment of ITS from small real-time data to big data mean that it now includes vast quantities of historical data as well. With these changes in data environments, advanced data management systems (ADMS) have recently been introduced to process, store, and analyze big data in the field of ITS.

The forecasting of traffic conditions along road networks is one of the essential factors with regard to ITS, as the social benefits of ITS can be maximized by proactive traffic management and the provision of dynamic traffic information based on the forecasting of traffic conditions. The traffic state forecasting system is one of the kernel sub-systems of an advanced traffic management system and the advanced traffic information system in ITS. Hence, forecasting technology to generate future states is closely related to increments in ITS benefits and to the performance of the traffic state forecasting system. Various forecasting models, from simple and conventional to refined and sophisticated, have therefore been proposed since ITSs were widely introduced at the end of the twentieth century.

Although existing advanced models have essentially achieved the common goal of ITS forecasting with improvements in forecasting accuracy, they have several chronic or emerging problems to be solved. First, the temporal prediction horizon of the models in most cases still operate on the short term; they cannot from the perspective of long-term forecasting overcome uncertainties in future states, and this remains an unsolved problem in the ITS forecasting area. Second, their structures are not suitable when they are coupled with a data management system such as ADMS and then used to estimate future states using both real-time and historical data, as they are in many cases designed to utilize only real-time data. Lastly, many sophisticated models becoming associated with obstacles which require field staff to manage. These models inevitably require the field staff to possess a deep understanding of the behaviors of traffic flows and forecasting model and then to manipulate the operational factors of these entities, such as structural changes of algorithms, in-and-out alterations of data, recalibrations of parameter values and other such actions.

In this thesis, a traffic information forecaster termed the KJC forecaster is developed based on a combination of k -nearest neighbor nonparametric regression and j -clustering using both real-time and historical data. First, a conquest solution to address the uncertainties of future states is proposed. In order to reduce the uncertainties of future states, the spatial concept of a forecasting model is expanded from an isolated location or link to a road network, after which traffic states, such as link travel speeds and link probe volumes, of the road network are utilized as the inputs to the forecaster. Second, KJC forecaster is designed to be used in conjunction with data management systems such as ADMS and to estimate the future traffic conditions of road networks. This can in turn be used to generate short-, middle-, and long-distance path travel times. The forecaster, therefore, is at the very least suitable for more proactive and tactical advanced traffic management and especially for dynamic intercity path travel times. The KJC

forecaster consists of three modules: a knowledge discovery module to search for and compile the information on future traffic states included in the historical data, a clustering module to determine decision-making groups, and a forecasting decision-making module which is based on case-based reasoning. The three modules were developed while considering operational requirements such as multivariate in-and-out data, easy alterations of inputs and outputs, the automatic processing of missing data, the automatic calibration of parameter values, as well as high-speed computing that is actually faster than real-time.

The performance of the traffic information forecaster was tested under the circumstances of large-scale data and the test bed was Seoul-Daejeon road section, 142 km, of Gyeongbu motorway. The historical database used, with a data size of 18,768,960 items, was composed of link-based traffic flow information and path travel times which were compiled using nearly four hundred million instances of trip chain data for eight months. This data was collected by means of dedicated short range communications technology. The results show that KJC forecaster estimates accurate traffic conditions, at least from a forecasting perspective, of road networks up to six hours in the future, with high-speed computations. In addition, the forecaster is clearly superior to two compared path travel time methods, instantaneous and experience-based, in terms of prediction accuracy of the path travel time. Therefore, it is clear that KJC forecaster as proposed in this thesis is a promising multivariate long-term traffic flow forecasting approach which is feasible for use with large-scale data.

***keywords* : Advanced Data Management System, Large-scale Data, k -Nearest Neighbor Nonparametric Regression, j -Clustering, Long-term forecasting, Traffic Condition of Road Network, Intercity Path Travel Time**
***Student Number* : 2008-30675**